

**RACORO Continental Boundary Layer Cloud Investigations. Part I:
Case Study Development and Ensemble Large-Scale Forcings**

Andrew M. Vogelmann¹, Ann M. Fridlind², Tami Toto¹,
Satoshi Endo¹, Wuyin Lin¹, Jian Wang¹, Sha Feng^{3,4}, Yunyan Zhang⁵, David D. Turner⁶,
Yangang Liu¹, Zhijin Li^{4,3}, Shaocheng Xie⁵, Andrew S. Ackerman²,
Minghua Zhang⁷, and Marat Khairoutdinov⁷

¹*Brookhaven National Laboratory*

²*NASA Goddard Institute for Space Studies*

³*UCLA Joint Institute for Regional Earth System Science and Engineering (JIFRESSE)*

⁴*Jet Propulsion Laboratory, California Institute of Technology*

⁵*Lawrence Livermore National Laboratory*

⁶*NOAA National Severe Storms Laboratory*

⁷*Stony Brook University*

Accepted for publication in

Journal of Geophysical Research – Atmosphere

Corresponding author: Andrew Vogelmann, Brookhaven National Laboratory, Biological,
Environmental & Climate Sciences Department, Bldg 490D, Upton, NY, 11973. Tel: 631-344-
4421. E-mail: vogelmann@bnl.gov

ABSTRACT

Observation-based modeling case studies of continental boundary layer clouds have been developed to study cloudy boundary layers, aerosol influences upon them, and their representation in cloud- and global-scale models. Three 60-hour case study periods span the temporal evolution of cumulus, stratiform, and drizzling boundary layer cloud systems, representing mixed and transitional states rather than idealized or canonical cases. Based on in-situ measurements from the RACORO field campaign and remote-sensing observations, the cases are designed with a modular configuration to simplify use in large-eddy simulations (LES) and single-column models. Aircraft measurements of aerosol number size distribution are fit to lognormal functions for concise representation in models. Values of the aerosol hygroscopicity parameter, κ , are derived from observations to be ~ 0.10 , which are lower than the 0.3 typical over continents and suggestive of a large aerosol organic fraction. Ensemble large-scale forcing datasets are derived from the ARM variational analysis, ECMWF forecasts, and a multi-scale data assimilation system. The forcings are assessed through comparison of measured bulk atmospheric and cloud properties to those computed in 'trial' large-eddy simulations, where more efficient run times are enabled through modest reductions in grid resolution and domain size compared to the full-sized LES grid. Simulations capture many of the general features observed, but the state-of-the-art forcings were limited at representing details of cloud onset, and tight gradients and high-resolution transients of importance. Methods for improving the initial conditions and forcings are discussed. The cases developed are available to the general modeling community for studying continental boundary clouds.

1. Introduction

Continental boundary layer clouds are important to simulations of weather and climate because of their impact on surface energy and moisture budgets, which are closely linked to moist convection and the atmospheric boundary layer structures and turbulence profiles [e.g., *Slingo et al.*, 2004; *Kollias and Albrecht*, 2000; *Ghate et al.*, 2010]. Shallow cumulus convection regulates the surface radiation budget [e.g., *Berg et al.*, 2011] and contributes to the preconditioning of deeper convection [*Chaboureau et al.*, 2004; *Khairoutdinov and Randall*, 2006; *Rio et al.*, 2009; *Zhang and Klein*, 2010]. However, the understanding and representation of boundary layer clouds in forecast and General Circulation Models (GCMs) has been a challenge. Model-parameterized boundary layer clouds do not agree well with observations at least in part because small-scale turbulence and convection are not properly represented [e.g., *Lenderink et al.*, 2004]. GCMs commonly misrepresent the diurnal cycle of continental convection, with rain onset typically occurring too early in the day [*Betts and Jakob*, 2002, *Bechtold et al.*, 2004; *Zhang and Klein*, 2010], which partly stems from parameterization deficiencies in cumulus cloud-top detrainment to the lower troposphere [*Guichard et al.*, 2004; *Rochetin et al.*, 2014].

Historically, a large majority of the literature on warm boundary layer cloud processes has focused on maritime clouds with comparatively few studies devoted to continental clouds [*Del Genio and Wolf*, 2000; *Mechem et al.*, 2010; *Ghate et al.*, 2010; *Zhang and Klein*, 2013]. This focus is likely attributable to the substantial role that marine boundary layer (MBL) clouds play in global climate sensitivity [*Hartmann et al.*, 1992; *Bony and Dufresne*, 2005], and also because weather forecasting over land primarily focuses on precipitation and severe weather [*Ghate et al.*, 2010]. For example, the Global Atmospheric System Studies (GASS; formerly GCSS)

boundary layer group has been instrumental in developing case studies for use by the general modeling community. However, the boundary layer cases focus on MBL clouds, where only one of the eleven past GASS boundary layer cases is a continental case [Brown *et al.*, 2002]. The MBL cases do not test model physics under the strong diurnal variability and range of aerosol loadings associated with continental boundary layer clouds, which would likely benefit MBL understanding. Although continental cloud is more transient than maritime cloud with obvious differences in underlying surface characteristics, much of the understanding of the two systems and questions about their formation and decay can be shared [Del Genio and Wolf, 2000]. For example, similar to the model mixing issues cited above for continental clouds, the largest source of uncertainty in cloud feedback may stem from the simulated strength of convective mixing between the lower and middle atmosphere in the tropics [Zhang *et al.*, 2013; Sherwood *et al.*, 2014]. Further, biases in simulated downward surface radiation are found to be similar for continental and maritime boundary layer clouds. Specifically, the same opposite-and-partially-compensating biases are found in forecasts from the European Centre for Medium-Range Weather Forecast (ECMWF): the broken cloud regime, typical of trade cumulus, exhibits liquid water paths and reflectivities that are too high, whereas the overcast regime, typical of stratocumulus, exhibits liquid water paths and reflectivities that are too low [Ahlgrimm and Forbes, 2012, 2014]. Thus, there is broad benefit for a greater variety of continental boundary layer cloud case studies.

The GCSS continental case available, Brown *et al.* [2002], is a 14.5-hour daytime simulation for fair-weather cumulus, which uses a modified initial sounding and surface flux time series from the Southern Great Plains (SGP) Site of the Atmospheric Radiation Measurement (ARM) Climate Research Facility [Stokes and Schwartz, 1994; Ackerman and Stokes, 2003; Mather and

Voyles, 2013]. The observed conditions were simplified, for instance by fixing geostrophic winds to be 10 m s^{-1} westerlies throughout the simulation; conditions were also modified to improve model performance, for instance by making a substantial adjustment to the initial sounding to offset apparent errors in the derived large-scale forcings as described in *Brown et al.* [2002; *e.g.*, see their Figure 1]. After case development, *Brown et al.* [2002] report that the impact of the large-scale forcings imposed on the models was negligible, indicating that this case can be viewed as a relatively canonical case of fair-weather cumulus. Such canonical case studies simplify interpretation and provide valuable benchmarks, but do not offer a means to test the wider range of cloud variability and transitions that occur in the atmosphere.

Long-term observations from ground-based remote sensing programs such as ARM and Cloudnet [*Illingworth et al.*, 2007] provide detailed statistics of clouds and their environment that span the full range of conditions at a single site. For example, the ARM SGP Facility has provided 20 years of data for statistical studies of boundary layer turbulence [*e.g.*, *Turner et al.*, 2014b] and cloud properties and radiative impacts [*e.g.*, *Dong et al.*, 2005, 2006; *Berg and Kassianov*, 2008; *Berg et al.*, 2011], drizzle statistics [*e.g.*, *Kollias et al.*, 2007], and vertical velocity statistics [*e.g.*, *Ghate et al.*, 2010; *Chandra et al.*, 2013]. Using over a decade of SGP data, a climatology was developed for systematic assessment of shallow-to-deep convective transitions [*Zhang and Klein*, 2010], and to examine differences between fair-weather cumuli that are forced and active [*Zhang and Klein*, 2013]. These studies are based on measurements within a narrow column over the SGP facility. Recently, ARM deployed scanning cloud radars that enable tracking the evolution of shallow cumulus and their properties, such as in the pioneering study by *Borque et al.* [2014].

However, even with the major advances in surface-based remote sensing, in-situ aircraft sampling is needed to provide detailed aerosol, cloud microphysical, and dynamical properties that cannot yet be retrieved. This is particularly true for warm boundary layer clouds, which often have liquid water paths (LWPs) less than 100 g m^{-2} that are referred to as Clouds with Low Optical Water Depths (CLOWD) [Turner *et al.*, 2007b]. CLOWDs challenge the limits of commonly used remote sensing techniques, which is particularly problematic since the Earth's radiative energy balance is very sensitive to small perturbations in LWP when LWP is small [Sengupta *et al.*, 2003; Turner *et al.*, 2007b]. To obtain detailed in-situ observations under CLOWD conditions, the ARM Aerial Facility (AAF) [Schmid *et al.*, 2014] conducted the Routine AAF CLOWD Optical Radiative Observations (RACORO) campaign [Vogelmann *et al.*, 2012], a first-of-its-kind, extended-term cloud aircraft campaign. RACORO operated for five months over the SGP, from 22 January to 30 June 2009, with the objective of obtaining a comprehensive in-situ statistical characterization of boundary layer clouds and aerosols for use in process studies, fine-scale model evaluation, and refinement of retrieval algorithms. The data has been valuable for observation-based studies of entrainment [Lu *et al.*, 2012a,b; 2013; 2014] and validation of retrieved water vapor turbulence profiles [Turner *et al.*, 2014a]. Modeling studies, however, require analysis and synthesis of comprehensive datasets for model initialization and evaluation, which involves a great amount of work.

As part of the multi-institution FAsT-physics System TEstbed and Research (FASTER) Project, here we document newly developed modeling case studies based on observations from the RACORO Campaign and the ARM SGP site for use in studies of continental cloudy boundary layers, aerosol influences upon them, and their representation in cloud- and global-scale models. The RACORO payload flown provides comprehensive in-situ measurements of

cloud, aerosol, radiation, and atmospheric state parameters that, combined with the extensive SGP measurements, provide multiple observational constraints on case study conditions. The wealth of multi-instrument aerosol and cloud data used here is uncommon for case studies, save for exceptions such as those from the Indirect and Semi-Direct Aerosol Campaign (ISDAC) [McFarquhar *et al.*, 2011]. Our goal is to use observations to derive multiple, realistic modeling case studies with a diversity of shallow continental cloud conditions that include mixed or transitional states that may not be considered in purely canonical cases. This is intended to enable testing model simulations over the range of processes that govern the cloud lifecycle—generation to dissipation—in a way that idealized cases may not.

This paper is the first in a three-part series that provides an end-to-end analysis of boundary layer cloud processes in observations and models, ranging from in-situ and surface-based observations to large-eddy simulations (LES) to single-column model (SCM) diagnostics. This first part focuses on case study generation, model aerosol specification, ensemble large-scale forcings, and evaluation metrics. The cases are designed to have a standardized modular configuration for ease of use in LES and SCMs alike. In part II, *Endo et al.* [Endo, S., A. M. Fridlind, W. Lin, A. M. Vogelmann, T. Toto I, A. S. Ackerman, G. M. McFarquhar, R. C. Jackson, and Y. Liu, RACORO Continental Boundary Layer Cloud Investigations. Part II: Large-Eddy Simulations of Cumulus Clouds and Evaluation with In-Situ and Ground-Based Observations, submitted to *J. Geophys. Res.*, 2014; hereafter referred to as *Endo_PartII*] use LES and in-situ microphysics observations for a cumulus-dominated case to assess bulk and bin microphysics simulations and their sensitivity to treatments of supersaturation and aerosol. In part III, *Lin et al.* [Lin, W., Y. Liu, A. M. Vogelmann, A. Fridlind, S. Endo, H. Song, S. Feng, T. Toto, Z. Li, and M. Zhang, RACORO Continental Boundary Layer Cloud Investigations. Part III:

Separation of Parameterization Biases in Single-Column Model CAM5 Simulations of Shallow Cumulus, submitted to *J. Geophys. Res.*, 2014; hereafter referred to as *Lin_PartIII*] use the LES runs in part II and in-situ dynamics observations to diagnose the sources of biases in the shallow convection simulated by an SCM for the same cumulus-dominated case.

This paper is organized as follows. Section 2 introduces case studies selected from the RACORO period, including their diversity and representativeness of the overall population of candidates. In-situ physical aerosol measurements are used in section 3 to develop a detailed aerosol model specification of size distribution and hygroscopicity parameter uncommon in case studies. In section 4, we describe an ensemble of large-scale forcing datasets to be used. The forcings are assessed using LES runs in a ‘trial’ configuration, described in section 5, where more efficient run times are enabled through minor reductions in grid resolution and domain size compared to the full-sized LES grid. Section 6 describes the data used for model constraints and evaluation. In section 7, the simulations are used to examine the forcings and their sensitivity to factors such as initial conditions and relaxation. The summary and conclusions are given in section 8, including a discussion of how the methodology used here, particularly with respect to initial conditions, could be improved.

2. Case Studies from the RACORO Field Campaign

In this section, the RACORO campaign and criteria used to select modeling case studies are discussed. The selected cases are then described and their representativeness appraised.

2.1. The RACORO Aircraft Field Campaign

During RACORO [Vogelmann *et al.*, 2012], the Center for Interdisciplinary Remotely-Piloted Aircraft Studies (CIRPAS) Twin Otter collected 260 h of data during 59 research flights in the vicinity of the ARM SGP Site, near Lamont, Oklahoma. Of these flights, 31 were cloud

flights that mostly sampled cumulus and stratocumulus, and 28 were non-cloud flights to characterize boundary layer properties such as turbulence, aerosol properties, and surface albedo. The cloud flights were designed to obtain an unbiased sampling of the cloud-field properties by flying a fixed pattern rather than seeking clouds with desired properties that could be statistically unrepresentative of the cloud field.

For each flight, the CIRPAS Twin Otter made comprehensive measurements valuable for case study development and model evaluation metrics (see Appendix A for a summary of the measurements). An overview of the diversity of continental boundary layer cloud fields sampled is given in Figure 1 in terms of (a) microscale and (b) macroscale properties. A scatter plot for each surveys the joint distributions between two key parameters measured during the flight periods. The microscale plot (Fig. 1a) is in terms of in-situ cloud-droplet number concentration (N_d) and boundary layer CCN at 0.2% supersaturation, $CCN_{0.2\%}$. N_d is a sum over the measured in-situ drop size distribution from the Cloud and Aerosol Spectrometer (CAS) for diameters $> 2 \mu\text{m}$, requiring that the liquid water content (LWC) calculated from those particles is $\geq 0.01 \text{ g m}^{-3}$. Figure 1a shows that a wide range of states are present, wherein the lowest N_d tend to coincide with stratus and the greatest N_d with cumulus. The macroscale plot (Fig. 1b) is in terms of cloud fraction and cloud LWP. Cloud fraction is the cloud frequency of occurrence per 20-min interval derived from the ARM Active Remote Sensing of Clouds (ARSCL) product [Clothiaux *et al.*, 2000], which uses radar and lidar measurements to generate a vertically resolved cloud mask for the narrow column above the instruments (see Appendix B). Here, we do not follow the common practice of using LWP measured by the surface-based microwave radiometers (MWR) because they have an RMS uncertainty of 20-30 g m^{-2} [Turner *et al.*, 2007a] that is relatively large for boundary layer CLOUDs whose LWP $< 100 \text{ g m}^{-2}$. Instead, in the

absence of a better approach, we estimate LWP from the product of PDFs of in-situ LWC from the CAS and cloud thickness from ARSCL. These estimates are used in Figure 1b showing that cloud fraction ranges from overcast to nearly clear, and that all median LWPs are less than the 100 g m⁻² definition for CLOUDs. Note that ARSCL cloud-top height for these clouds was predominately determined by the lidar, rather than the radar that may not detect cloud with small drops and low LWC. The ARSCL-obtained cloud thicknesses agree well with the range from *Vogelmann et al.* [2012, Figure 8], who showed that bulk LWP statistics from an MWR (where RMS is largely averaged out) agree well with the LWP computed from in-situ CAS LWC for cloud thicknesses of 500 m \pm 200 m (*i.e.*, \pm 40%). Note that when LWP < 50-75 g m⁻², which is greater than many of the values in Figure 1b, the 40% uncertainty in cloud thickness incurs an LWP uncertainty less than the 20-30 g m⁻² MWR uncertainty. A regression between the LWP product in Figure 1b and LWP retrieved by the MWR finds good correlation on average, but with an expectedly large RMS (not shown).

Thus, multiple possibilities exist for case study generation. Owing to the manually intensive nature of creating case studies, only a limited number of flight days can be used here. The selection criteria and cases selected are described next.

2.2. Case Study Selection and Descriptions

A three-pronged observation-LES-SCM approach was used for case selection. The most important consideration was from the in situ observations, where we sought consecutive multi-day periods that had aircraft flights with the greatest amount of in-cloud sampling (for a given cloud type). Multi-day periods were preferred to include simulation of shallow cloud that persist during the night or formed in early morning hours in a realistically spun-up state. Variability of cloud type between flight days was preferred when possible, for which updraft strength and

large-scale boundary layer and/or environmental factors often varied, such as CCN. Multiple periods were viable based on these considerations. Three 60-h cases were selected as the subset that captures the best diversity within a limited number of cases. Before finalizing the selections, checks were conducted with preliminary large-eddy simulations to confirm that large-scale forcings are of sufficient quality for viable simulations. Close examination of these forcings is the subject of section 7. Finally, SCMs, run for the entire RACORO period, were used to verify that a candidate case had the characteristic errors in its cloud simulation. Unlike the previous tests, it was used to reassure that our selections could be useful to diagnose GCM shortcomings. Since all of the initial cases tested were problematic, this last check did not challenge our selected cases. Simulations were conducted using the FASTER SCM Testbed for three SCMs from U.S. modeling centers: the National Center for Atmospheric Research (NCAR) Community Atmosphere Model (CAM) version 5 (CAM5), which is the atmospheric model of the Community Earth System Model (CESM) [Neale *et al.*, 2012]; the Geophysical Fluid Dynamics Laboratory (GFDL) Atmosphere Model (AM) version 3 [Donner *et al.*, 2011]; and the NASA Goddard Institute for Space Studies (GISS) ModelE [Schmidt *et al.*, 2006]. The results confirmed that multiple SCMs, given the same large-scale forcing, have different issues representing the low-level clouds in the case studies (not shown).

The general characteristics of the three 60-h cases are discussed next and further details of their time evolution are described in section 7.1. In the discussions that follow, time is expressed as decimal day in May [UTC].

a) Case 1. Cumulus with variable aerosol, 22-24 May: This case is dominated by active fair-weather cumulus. The large-scale state is steady, at least with respect to the cloud regime. Cloud is locally generated in sync with the diurnal cycle of surface heat fluxes driven by

solar heating. Figure 2a shows that a cumulus cycle repeats over the three days, wherein cumulus are generated mid-morning, and the cloud-base heights increase by 0.5 to 1 km over the day before dissipating in mid-to-late afternoon. The median cumulus updrafts from 1-Hz in-situ measurements are $\sim 1 \text{ m s}^{-1}$, where cloud periods are defined as those with CAS LWC $\geq 0.01 \text{ g m}^{-3}$ (see section 2.1). (The exact locations of the updrafts relative to the cloud vertical boundaries are not well known because the clouds sampled were relatively thin and resided within a boundary layer that varied spatially and temporally.) Over the period, median $\text{CCN}_{0.2\%}$ drops by almost a factor of two, from 600 to 350 cm^{-3} , providing some range in aerosol loading.

b) Case 2. Cumulus and drizzling Stratus, 26-28 May: This case represents a transition from cumulus to drizzling stratus (day 1 through day 2) to essentially clear skies (day 3) (Figure 2b). The stratus follows a weak cold-front passage at 27.5 May, with steady drying occurring afterwards. There was not an aircraft flight on 28 May, the clear-sky day at the end of the period. It is included in the case to test whether SCMs and LES can properly simulate the limited onset of daytime cloud. For a follow-on project, the clear-sky period is also used to test models using SGP Raman lidar profiles of boundary layer state variables (discussed in section 7.1), and retrievals of boundary layer turbulent fluctuations in water vapor mixing ratio [Turner *et al.*, 2014a]. Over the first two days, median in-flight updrafts decreased by a factor of nine, from 0.9 to 0.1 m s^{-1} , and median $\text{CCN}_{0.2\%}$ were consistently among the lowest of the three case studies, being $170\text{-}280 \text{ cm}^{-3}$.

c) Case 3. Variable Cloud Types, 6-8 May: This case poses the greatest challenge to models and to case study development with air mass changes that resulted in multiple cloud-type transitions, from stratus to stratocumulus (day 1), from stratus to cumulus (day 2), and finally

to stratocumulus following nocturnal convection (day 3) (Figure 2c). An occluded front, which had been south of the SGP, developed into a warm front and moved over the site at 7.6 May, followed by a southerly flow of warm, moist air from the Gulf. At 8.5 May, a slow-moving cold front passed overhead as a squall line passed to the north of the central facility, which was followed by a drying pattern. The median updrafts over the three days were low and varied, respectively 0.4, 0.8, 0.4 m s⁻¹. The median CCN_{0.2%} was quite variable with day 2 being triple that for the other days, at 200, 580, and 215 cm⁻³.

Figure 1 shows that these cases are generally representative of the overall population during RACORO. The selected days are denoted by the dates given below their medians, and the case number is denoted by the symbols encircling the date. Figure 1a shows that the selected days reside across the range of the N_d -CCN_{0.2%} points, although not capturing extreme N_d and CCN_{0.2%} values. Figure 1b shows that the selected days and their percentiles are also representative of the overall population in terms of LWP and cloud fraction.

3. Model Aerosol Treatment

These case studies benefit from the availability of aircraft measurements for detailed specification of aerosol number-size distribution profiles and an aerosol hygroscopicity parameter. Combined, these parameters constrain aerosol properties relevant to droplet activation and aerosol-cloud interactions that are commonly neglected or highly simplified in modeling case studies.

3.1. Aerosol Size Distribution Profiles

Profiles of aerosol size distribution were generated from measurements during aircraft spirals, performed at the beginning and end of each flight, made by a Scanning Mobility Particle Sizer (SMPS) and a Passive Cavity Aerosol Spectrometer Probe (PCASP). Each spiral ranged

from 150 m above-ground level (AGL) to about 450 m above cloud top. The size distribution measurements were binned into 100-m height intervals and cloud screened (see Appendix C). The measured dry aerosol size distributions for each 100-m interval were fit using up to three lognormal functions for use in models,

$$\frac{dN}{d \log_{10} D_p} = \sum_{i=1}^3 \frac{N_i}{\sqrt{2\pi} \log_{10} \sigma_i} \exp \left[-\frac{1}{2} \left(\frac{\log_{10} D_p - \log_{10} D_{p,i}}{\log_{10} \sigma_i} \right)^2 \right], \quad (1)$$

where, for each mode i , N_i , $D_{p,i}$ and σ_i are, respectively, the total number concentration, particle geometric mean diameter, and geometric standard deviation. (Note that the lognormal distribution is commonly defined in terms of natural logarithms but we prefer to cast it in terms of base-10 logarithms; the parameters are easily converted between the two.) For ease of use by modelers, a simplified version of the profile fits was generated, where the geometric mean radius and standard deviation are fixed per mode for a given profile and only the total number concentrations per mode vary with altitude [Fridlind *et al.*, 2012]. The fixed mean radius and standard deviation per profile were determined from the 100-m interval fits using a number-concentration-weighted average of their values through the column, which favors the higher concentrations below cloud that are most important for cloud droplet activation. Given the fixed values, the number concentrations per mode were computed for each height interval. See Figure 3 for a representative example of the fitting. The total aerosol counts from the fitted aerosol size distributions were verified to be in excellent agreement with the counts from a condensation particle counter with a lower size cutoff of 15 nm (not shown). Although aerosols in the smallest-sized mode are not generally large enough to be activated, it was considered most straightforward to fit the full aerosol size range with overlapping modes and avoid imposing case-dependent assumptions about activation size as a function of hygroscopicity.

Thus, the aerosol size distribution profiles are available in three formats: (1) raw data, consisting of single size distribution per height interval from the combined SMPS and PCASP data; (2) an intermediate format, consisting of trimodal fits of (1); and (3) a simplified profile, where the geometric mean radius and standard deviation are fixed and the number concentrations vary with altitude. In this study, we use format (3) only. Note that, as described in section 5.4, the three modes are interpolated in time for use in the simulations. In practice, owing to the intermittent presence of prominent peaks across a wide size spectrum, four modes were used for each case, wherein no more than three modes contained particles at any single time.

3.2. Aerosol Hygroscopicity

Aerosol hygroscopicity is quantified from aircraft aerosol measurements in terms of a single parameter, κ [Petters and Kreidenweis, 2007; Moore et al., 2008; Ghan et al., 2011a,b]. The value of κ is zero for insoluble materials such as soot or mineral dust, ~ 0.1 for secondary organic aerosols [e.g., Wang et al., 2008; Mei et al., 2013a, b], ~ 0.55 for ammonium bisulfate, and ~ 1 for sea salt [Petters and Kreidenweis, 2007; Rose et al., 2010]. For particles of a chemically diverse mixture of soluble and insoluble components, the overall κ for the particles is simply the volume average of the κ for participating components [Petters and Kreidenweis, 2007]. Values of κ for continental and marine aerosols typically cluster around 0.3 and 0.7, respectively [Andreae and Rosenfeld, 2008; Rose et al., 2010].

A single κ value is derived for the aerosol population from each spiral based on the measured aerosol size distributions and CCN spectra (see Appendix D for details). The procedure is tested by comparing the observed CCN spectra to that computed from the derived κ and height-averaged size distributions per spiral. An example for the two spirals on 22 May is given in Figure 4. As is generally the case, the comparisons agree well, particularly for the lower

supersaturations. Note that the CCN standard error at 0.2% supersaturation is imperceptibly small since it benefits from the combined sampling of the scanning 0.2% values and the column fixed at 0.2%. Overall, the fits for the three cases agree very well with the observations, assuring that the aerosol representations (size distribution and κ) faithfully represent the CCN spectra.

The κ values obtained are notably low. For the example in Figure 4, the κ for spiral 1 and 2 are, respectively, 0.08 and 0.07. Similarly low values were found for all cases, as shown in Table 1. The values are about ~ 0.10 , with a minimum of 0.04 and a maximum of 0.13. These values are somewhat less than the ~ 0.3 typically observed at continental sites, suggesting a larger contribution from species with low hygroscopicity such as organics and/or black carbon during our cases. For example, such low κ values have also been found for heavily forested regions, such as the Amazon [*e.g.*, Pringle *et al.*, 2010] that has a large aerosol organic fraction. Unfortunately, hygroscopicity and chemical composition of surface aerosol at the SGP site did not begin until a year after the RACORO field campaign when operations started for an Aerosol Chemical Speciation Monitor (ACSM) and Tandem Differential Mobility Analyzer (TDMA). Calculation of size-segregated κ from TDMA data collected during May of 2011 for another project indicate values commonly spanning 0.05–0.15 (not shown). Also consistent with our finding, analysis of ACSM data also indicate a large organic fraction is present during May [Parworth *et al.*, 2015]. The implication of these low κ values in cumulus simulations is explored by *Endo_PartII*.

4. Ensemble Large-Scale Forcing Data

The simulations are driven by large-scale forcings, defined here as large-scale vertical wind and horizontal flux divergence profiles for heat and moisture over a domain. In configuring modeling case studies focused on simulating shallow clouds with LES, large-scale forcings are

generally simplified and tuned simultaneously in a manner that might offset other case study or model deficiencies, be it in initial conditions, model physics, dynamics, or other factors [e.g., Jiang *et al.*, 2000]. Here we avoid any tuning of observation-based inputs by, instead, taking an approach intended to bracket uncertainties in input or forcing terms by using multiple sets of large-scale forcings, collectively referred to as ensemble forcings. Three independent sets of forcing data are prepared for the RACORO case studies, with each set available for the standard ARM forcing SGP domain (~300 km by 300 km) and for a reduced domain (150 km by 150 km). All data sets are hourly and, in addition to large-scale vertical wind and flux divergence profiles, include domain-mean profiles of temperature (T), water vapor mixing ratio (q_v), horizontal wind components (u , v), domain-mean surface fluxes, and other surface and near surface variables.

The ‘ARM’ large-scale forcing data are based on the constrained variational analysis approach described in Zhang and Lin [1997] and Zhang *et al.* [2001]. The forcing for the SGP standard domain (300 km) is part of the ARM continuous forcing data [Xie *et al.*, 2004] that has been extensively used and tested within ARM. Upper atmospheric data are from the National Oceanic and Atmospheric Administration (NOAA) Rapid Update Cycle (RUC) analysis, which is constrained with ARM surface observations and NASA satellite measurements to produce the forcing data. The data are produced at constant pressure levels with a 25-hPa vertical grid spacing. The forcing for the reduced domain (150 km) is specially produced for the RACORO-FASTER modeling study. The constrained variational analysis was applied to adjust the RUC analysis over the reduced domain, and the forcing data are available at a higher vertical resolution of 10 hPa.

The ‘ECMWF’ forcing data are derived from the ECMWF short-term 12–35 hour forecasts that are initialized daily at 12 UTC (*i.e.*, forcing data for each day [0-23 UTC] is the 12-35 h

forecast initialized at 12 UTC of the previous day). The data over the ARM sites are archived by the ARM External Data Center, and are used here to generate a complete set of forcing data. The source data has 91 hybrid sigma-pressure vertical levels and 0.225° horizontal resolution (~ 22 km at the SGP). During the RACORO period, the forecasts at the SGP had 20 levels below 700 hPa with a vertical spacing gradually increasing from 3 hPa near the surface to 25 hPa by 700 hPa. The derived forcing data are averages over the grid cells within both the SGP standard and reduced domains, and both use the same high-resolution vertical grid. (We note that the LES performance here might differ from the ECMWF forecasts that use their parameterizations and also have the large-scale forcings at a higher grid resolution.)

Lastly, the ‘MS-DA’ forcing data are derived from the WRF-3DVar-based multi-scale data assimilation (MS-DA) system developed under the FASTER project [Li *et al.*, 2012; Li *et al.*, 2014]. A three-domain nested WRF configuration is used, centered on the SGP central facility, where the innermost domain has a 2-km grid spacing and an area that closely matches the ARM standard domain. The three domains have 45 vertical layers with the top at 100 hPa. The model is driven by North American Regional Reanalysis (NARR), followed by an MS-DA algorithm that decomposes the cost-function minimization process to effectively assimilate high-resolution observational data to produce a fine-resolution analysis that is used to derive large-scale forcing data over the domain of interest. The model integration and fine-scale data assimilation cycle used here are the same as in Li *et al.* [2014]. Here we use only the large-scale vertical wind and horizontal heat and moisture flux divergence profiles, but it is worth noting that the MS-DA data can also provide scale-aware forcings [Feng *et al.*, 2015]. The MS-DA forcing data for the standard domain has the same pressure levels as the ARM standard forcing with a 25-hPa

vertical spacing. The reduced domain is similarly derived except at the native WRF model levels, which has two more layers between 950 and 975 mb than the standard forcing.

Note that the three sources of forcing data are derived quite differently: ARM uses the RUC as the first guess and a constrained variational analysis approach, ECMWF uses forecasts based on the ECMWF model and its initialization procedure, and MS-DA is initialized with the NARR and assimilates high-resolution data into a nested WRF by using the NCEP Gridpoint Statistical Interpolation (GSI) method. Collectively, the three forcing data sets with two domains each constitute a large-scale forcing ensemble to address large-scale uncertainties due to model physics and analysis methods. This ensemble is different than the commonly used statistical ensemble, which is generated from perturbing the measurement inputs within the bounds of their uncertainties [*e.g.*, Davies *et al.*, 2013]. There are benefits to both approaches. The statistical ensemble focuses on the input variable uncertainties to characterize the envelope of forcing states possible from known sources of error; however, it does not consider uncertainties in model physics or derivation method. Our ensemble accounts for such uncertainties and thereby provides an ensemble of ‘best estimates’ rather than a statistical envelope, which casts a wider net in considering inevitable dependences on model physics in the derivation of large-scale forcing. This enables insights into the simulated features that are independent of forcing type; however, attribution of the differences to specific model physics is difficult-to-impossible.

5. Description of ‘Trial’ Large-Eddy Simulations

The previous two sections describe the derivation of aerosol properties and ensemble large-scale forcings for multiple case studies with a modular specification. Next we test a series of possible specifications for each case using ‘trial’ large-eddy simulations, where more efficient run times are enabled through minor reductions in grid resolution and domain size compared the

‘full-sized’ simulations used for follow-up detailed investigations [*e.g.*, *Endo_PartII*]. Trial simulations are found to effectively capture the macroscale behavior of full-sized simulations and, therefore, they can be compared with observations of macroscale cloud and atmospheric properties to evaluate ensemble large-scale forcing datasets and other choices in specification. The best matches to macroscale observations are candidates for full-sized simulations and more intensive comparison with microscale observations. For example, in part II of this series, selected specifications are used in full-sized simulations for two independent LES models and the simulated cloud microphysical properties are compared with in-situ observations [*Endo_PartII*]. We note that closer agreement of a given trial LES result to macroscale observations is *not* interpreted as proof that the specification is necessarily more correct, but only that a full-sized simulation is justified for more detailed comparison with microscale observations.

5.1. The DHARMA Model

Nearly all trial simulations use the 3D Distributed Hydrodynamic Aerosol and Radiative Modeling Application (DHARMA) code [*Ackerman et al.*, 1995]. It treats fluid dynamics using an LES model [*Stevens et al.*, 2002] with a dynamic Smagorinsky sub-gridscale scheme [*Kirkpatrick et al.*, 2006], cloud microphysics with a modified version of the *Morrison et al.* [2005] two-moment microphysics scheme [*Fridlind et al.*, 2011], and radiative transfer with a two-stream model [*Toon et al.*, 1989]. Dynamics and cloud microphysics are integrated with a default time step of 1.5 s, which is shortened if needed to maintain a flow Courant-Friedrichs-Lewy number no greater than 0.8.

Beyond the standard prognostic variables for the fluid dynamics and two-moment microphysics schemes (described within references cited above), the model configuration here

455 additionally uses a prognostic aerosol number variable for each of four independent aerosol
456 modes. Each of those aerosol number variables is the sum of unactivated aerosol number
457 concentration within a mode plus the number concentration of cloud droplets activated from that
458 mode. Thus, cloud droplet activation does not affect the aerosol number variable for a mode,
459 because that variable includes the number concentration of activated droplets from the mode.
460 The maximum number concentration of cloud droplets in a grid cell is the total of the four
461 aerosol number variables in the grid cell. Where there are no cloud droplets, each of the aerosol
462 number variables is equal to the total number of aerosol particles within each mode. The
463 prognostic aerosol number variables are subject to advection and mixing.

464 All simulations begin with a cloud-free domain that is initialized with the quad-modal
465 aerosol size distribution fits derived from observations as described earlier. The height-
466 dependent aerosol number concentration for each mode represents an independent variable. The
467 hygroscopicity parameter calculated for ammonium bisulfate is replaced with the observation-
468 based κ values, unless otherwise noted. (The simplified baseline assumption that all aerosol is
469 ammonium bisulfate is made on the basis that sulfate is a leading inorganic component of
470 boundary-layer aerosol, but is most likely neither fully neutralized nor fully acidic [*e.g.*, Walker
471 *et al.*, 2004].) In the two-moment cloud microphysics model, droplet activation proceeds via
472 calculation of a prognostic saturation excess following *Morrison and Grabowski* [2008]. The
473 supersaturation used for activation is taken as the minimum of the supersaturation over a time
474 step (from the semi-analytic solution given in their appendix), which compares favorably with
475 activation from the size-resolved microphysics scheme run with DHARMA. Droplets are
476 activated from each aerosol mode as described by *Abdul-Razzak et al.* [1998] and *Abdul-Razzak*

and Ghan [2000]. Simulations assume that the droplet size distribution relative dispersion is a function of LWC following *Geoffroy et al.* [2010].

Radiative transfer is computed independently for each column once every minute using the two-stream model. The ARM standard domain large-scale forcing data set is used to specify time-varying skin temperature. Surface albedo and emissivity are assumed to be 0.2 and unity, respectively. Clear-sky downwelling infrared fluxes at the top of the DHARMA grid are computed offline as described in section 5.4.

5.2. Trial Dimensions

The trial horizontal and vertical resolution and domain size are each reduced by about 25% for more efficient run times, compared to the full LES configuration used in *Endo_PartII*. The trial domain is 7.2 km by 7.2 km (full: 9.6 km), horizontal grid spacing is 100 m (full: 75 m), vertical spacing is 50 m below 5 km (full: 40 m) and stretched to the total domain height of 15 km with 120 levels total (full: 145), and the time step is 2 s (full: 1.5-1.8 s). These configurations yield a trial mesh of 72x72x120 with $dt=2$ s versus the full mesh of 128x128x145 with $dt=1.5-1.8$ s such that the trial simulations run in only one-fifth the time of a full simulation.

The ability of the trial configuration to capture the behavior of the full-sized LES was tested by running the same simulation (Case 1, ARM standard domain with relaxation [discussed next]). The trial simulation closely tracked the full-sized results in terms of the time variation of domain-averaged LWP, cloud fraction, and in-cloud droplet number concentration and cloud effective radius (not shown). This result demonstrates that trial simulations can reproduce the time-dependence of bulk properties for the environments under study. However, note that the configuration might not necessarily generalize to environments with stronger mesoscale flows

with large horizontal gradients across the LES domain, requiring separate study to determine appropriate configurations for those conditions.

5.3. Relaxation

Relaxation forcing (or nudging) may be used in multi-day simulations to offset the accumulation of error when large-scale conditions are not well known; it is perhaps most valuable when a leading objective is comparison of simulations with observations [e.g., *Neggers et al.*, 2012]. It is introduced by adding to the model state a relaxation term, $(X_{obs}-X_{model})/\tau$, where X_{obs} is the observed value, X_{model} is the model value, and τ is the relaxation timescale [e.g., *Randall and Cripe*, 1999]. When used in this study, relaxation is always applied to mean profiles to preserve horizontal heterogeneities [cf. *Fridlind et al.*, 2011].

Relaxation is always used in simulations to specify horizontal wind profiles. Profiles of the u and v wind components are relaxed with a three-hour time scale and a height-dependent strength of 0 for $z \leq 400$ m, 1 for $z \geq 600$ m, and a linear increase from 0 to 1 between 400 and 600 m. Relaxation is also always used to specify changes in aerosol profiles. The aerosol modal number concentration profile, geometric mean diameter, geometric standard deviation, and κ are relaxed to time-dependent observation-derived values with a six-hour time scale. Thermodynamic relaxation is not always used; when it is, q_v and θ are relaxed using a 12-hour time scale and the same height-dependent strength as for the winds.

The relaxation time scales applied (3-h winds, 6-h aerosol, 12-h thermodynamics) were chosen in keeping with the general range of values used within the community [e.g., *Fridlind et al.*, 2012; *Morrison et al.*, 2011; *Neggers et al.*, 2012]. Our values tend to be upper estimates of the ranges to apply as little influence as necessary. The values result from a compromise between keeping them short enough to hinder significant drift and capture day-to-day variability, but not

so short to operate at time scales on par with cloud lifetime. A sensitivity study was conducted for case 2 using shorter relaxation time scales of 1-h winds, 0-h aerosol (i.e., instantaneous), and 6-h thermodynamics. The effects (not shown) of the shorter time scales on the boundary layer and cloud properties studied here were found to be negligible except for thermodynamics, which showed a minor improvement in cloud duration.

We acknowledge that running LES with thermodynamic relaxation is inherently inconsistent with assessing forcing datasets. The relaxation term does not represent a real physical process [Randall and Cripe, 1999] and its effect inherently differs for each model depending on how far its simulation deviates from observations, which amounts to tuning the forcing to the model. We include relaxation here because the forcing datasets cannot be assumed accurate enough to produce simulated macroscopic cloud properties that are sufficiently similar to observations for the purposes of this study (even in a perfect model), particularly for our 60-h simulations (as demonstrated below). Sixty-hour simulations are preferred here (rather than restarting at the beginning of each day) to enable the model to simulate the full diurnal, multi-day evolution of the cloud systems, and to enable comparison of this evolution with SCMs. When relaxation is used, integrity should be maintained for processes that depend on the profiles of state variables, but large errors can result in processes that depend on rates (e.g., fluxes such as horizontal advection divergence) [Randall and Cripe, 1999]. To avoid undue influence of relaxation on cloud-scale processes, relatively long relaxation time scales are used, but they remain short enough to substantially hinder drift over 60-h periods (3-h winds, 6-h aerosol, 12-h thermodynamics). The primary aim of using relaxation here is to maintain simulation integrity in bulk atmospheric properties to enable the observation-informed study of finer processes, such as those that govern cloud microphysics and structure, but caution should be used if evaluating

boundary layer physics and dynamics. The objective in the next section is to assess the quality of the ensemble large-scale forcing datasets with and without relaxation, as well as other sensitivities to the simulation configuration.

5.4. Model Initialization and Configuration

A sounding near 12 UTC is used to initialize u , v , q_v , and potential temperature, θ , in each case. Initial random θ perturbations of 0.1 K are introduced below 500 m. Clear-sky downwelling infrared fluxes above model top at initiation are computed offline using the MODTRAN-v4 radiative transfer algorithm [Anderson *et al.*, 2001]. Its inputs are the sounding, with the column water vapor amount scaled to the microwave radiometer measurement, and the Modtran mid-latitude summer ozone profile scaled to the column measurement from the Ozone Monitoring Instrument. Surface momentum fluxes are obtained from the bottom-layer center using the Monin–Obukhov similarity as in Brown *et al.* [2002] for the previous GCSS SGP case, except the surface roughness length is modified from 0.035 m to 0.04 m. Given that roughness length is an important parameter for momentum flux [*e.g.*, Liu *et al.*, 2013], the modified value represents the average vegetation height typical in northern central Oklahoma during May [David Cook, *personal communication*]. Aerosol number concentration profiles in four lognormal modes with geometric mean diameter, geometric standard deviation, and κ are interpolated in time from per-flight mean values for each mode (number concentration varies with height, the other properties listed are uniform with height at any time).

Since our objective is to examine the sensitivity of simulations to different atmospheric advective components of the forcings, which are considered most uncertain since they are not directly observed (in contrast to initial soundings or surface turbulent fluxes), we isolate the advective effects by prescribing the other forcing components to a common reference. The

ARM standard domain forcing data is used as the reference for all simulations, providing the surface latent and sensible heat fluxes, and the skin temperature. When thermodynamic relaxation is used, simulations are always relaxed to the state profiles from the ARM standard domain forcing dataset. This approach of using specified values prevents the effects from interactions between surface temperature and fluxes. If they were allowed to interact, another complexity would be introduced by involving a land-surface model and its uncertainties. These RACORO cases can be developed later into coupled surface-atmosphere cases for study but, for now, we simplify the study to keep it tractable.

Another simplification is that the DHARMA ice physics is turned off and thus all clouds are treated as water clouds. This is appropriate since our focus is on warm boundary layer clouds and it eliminates any potential complications arising from the uncertainties of ice physics should clouds with subzero temperatures form. Such uncertainties include the treatment of mixed-phase clouds, and concentrations of ice nuclei and their initiation. Consequently, any portion of a simulation during or following a deep convective event is unreliable and will not be analyzed.

6. Model Evaluation Data

Simulations are evaluated with surface-based observations of domain-mean cloud macrophysical and boundary layer properties. The observations are briefly summarized here and details of the data metrics and their uncertainties are given in Appendix B.

- a. Liquid-water path. Two cloud LWP retrievals are used: the MWRRet microwave radiometer retrievals [Turner *et al.*, 2007a] that are routinely available with an uncertainty of 20–30 g m⁻²; and the Mixed-phase Cloud Retrieval Algorithm (MIXCRA) [Turner, 2007] infrared+microwave retrievals that are more accurate (*e.g.*, 30% for LWP < 5 g m⁻²) but available only for limited periods. For comparison to LES output, retrievals are time-

591 averaged to yield domain-averaged LWP, $\overline{\text{LWP}}$. Domain-averaged LWP is preferred to in-
592 cloud LWP (*i.e.*, $\overline{\text{LWP}}$ divided by areal cloud fraction) to avoid the need to define ‘cloud’ in
593 simulations and observations where not required.

594 b. Cloud fraction. Two estimates of cloud fraction are used to approximate the measurement
595 uncertainty. One is the hemispheric sky cover from the total-sky imager (TSI), and the other
596 is a column measurement of the ARSCL profile of cloud fractional occurrence per 10-min
597 interval. In model simulations, ‘cloud’ is defined as $\text{LWP} > 1 \text{ g m}^{-2}$, which is a rough
598 estimate of the lower detectability limit of the measurements [*e.g.*, MIXCRA].

599 c. Lifting condensation level height. The lifting condensation level (LCL) height is determined
600 from surface-air observations of relative humidity and temperature. Its domain-averaged
601 height, Z_{LCL} , and uncertainty are computed from the mean and standard deviation of the five
602 stations closest to, and including, the SGP central facility. The same calculation is applied to
603 model simulations for consistent comparisons.

604 d. Boundary layer moisture and temperature. Raman lidar measurements provide high
605 frequency vertical profiles ($\sim 75 \text{ m}$ every 10 min) of the boundary layer water vapor mixing
606 ratio, temperature, and relative humidity. Values 300-500 m above ground level (AGL) are
607 averaged using a one-hour moving window to produce mid-boundary-layer-averaged water
608 vapor mixing ratio, $\overline{q_v}$, temperature, \overline{T} , and relative humidity, $\overline{\text{RH}}$. (See Appendix B for
609 details.)

610 7. Results

611 The ensemble large-scale forcings and other choices in specification are assessed next. For
612 the simulation naming convention, the large-scale forcing type (ARM, ECMWF, MS-DA) is
613 followed by ‘D’ for standard-domain (300 km) forcings or ‘d’ for reduced domain (150 km)

forcings. A ‘+R’ denotes the use of 12-hour thermodynamic relaxation. For example, a simulation using ARM standard-domain large-scale forcings run with thermodynamic relaxation is ARM_D+R, and without relaxation it is ARM_D.

7.1. Observed boundary layer and cloud properties

The three cases are first described in terms of observed mid-boundary-layer-average (300-500 m) $\overline{q_v}$, \overline{T} , and \overline{RH} from the Raman lidar, and Z_{LCL} (black lines in Fig. 5) and cloud fraction profiles (Fig. 6, top row). For the case 1 cumulus (column 1), observations show a regular diurnal cycle with moderate amplitudes in $\overline{q_v}$ and \overline{T} of about 4 g kg⁻¹ and 5 K, and a \overline{RH} amplitude of 30-50% with negligible overall trend compared to other cases, indicating relatively stable mesoscale conditions. The diurnal cycle in Z_{LCL} is particularly pronounced with amplitudes of 1.5 to 2.0 km and a collapse to near-zero at nighttime. The cumuli follow a regular pattern (Fig. 6a), initiating ~10 local solar time (LST) and dissipating in the late afternoon about the time of peak Z_{LCL} (Fig. 5d). Note in Figure 6a that a cumulus cloud initiates at 22.6 May (7 LST); this is before solar heating and is likely in a residual layer of moisture from earlier events, thus forming earlier from a different mechanism than the main cumulus field that is coupled with surface fluxes.

For the case 2 cumulus and drizzling stratus (column 2), $\overline{q_v}$ steadily decreases over the simulation period while \overline{T} fluctuates with a minimum at the center of the period, both consistent with a gradual change of the air mass after a weak cold front passed about 27.0 May. The \overline{T} minimum coincides with 100% \overline{RH} (Fig. 5g) and follows generation of a post-frontal stratus deck around 27.1 May (Fig. 6f). The sharp dips in \overline{T} likely stem from below-cloud evaporative cooling during drizzle. The Z_{LCL} diurnal pattern is more irregular than in case 1, where Z_{LCL} is

suppressed to 1 km at 27.9 May, coinciding with the stratus coverage, followed by a large increase to 2.5 km the next day (29.0 May) under essentially clear skies (Fig. 6f).

For the case 3 variable cloud types (column 3), \bar{T} initially exhibits a minimum at 6.5 May, which coincides with 100% $\overline{\text{RH}}$ and likely is from evaporative cooling of drizzle, and then steadily increases to 295 K by 7.0 May. Peak Z_{LCL} at ~ 1 km are generally lower than for the previous two cases, associated with a reduction in surface fluxes caused by the daytime cloud cover for most of this period (Fig. 6k). Meanwhile, $\overline{q_v}$ undergoes a huge increase between 7.3 and 8.0 May, rising from 8 to 20 g kg⁻¹. At 7.0 May an occluded front was just south of the SGP that later developed into a warm front and moved over the site by 7.6 May. After it passed, a southerly flow developed that transported warm, moist Gulf air into the region (Figs. 5i,j). At 8.5 May, a slow-moving cold front passed, associated with reduced $\overline{q_v}$. However, a squall line passed to the north of the central facility at the same time as the cold front, which is associated with deeper convection in many trial simulations for which the applicability of only warm-cloud physics is problematic. For this reason, the simulations (but not the observations) after 8.5 May are blacked out.

In the analyses that follow, these observational time series are compared with simulated domain-average quantities. Spatial scales differ since the observational time series are derived from point measurements; however, for our application here, differences are considered insufficient to warrant generating ensembles of point properties from each simulation. Only multi-hour patterns are discussed (not higher frequency), consistent with averaging and smoothing applied to the observational data. As will be shown below, simulations often generate large deviations from observations that are well rendered with domain averages.

7.2. Sensitivity to advection and surface forcing components

The relative importance of sensible and latent heat fluxes and large-scale forcings to observed patterns are considered using the ARM_D forcing as a baseline and performing two sensitivity tests: surface fluxes as in ARM_D but with no large-scale forcings (ARM_D-noLS), and large-scale forcings as in ARM_D but with the domain-averaged surface fluxes replaced by values from only the SGP site central facility (ARM_D-SF). The domain-averaged surface fluxes in ARM_D are from Energy Balance Bowen Ratio (EBBR) measurements, and ARM_D-SF values are from an eddy correlation (ECOR) flux measurement system at the central facility. The former have far better spatial coverage (almost all over grassland), while the latter provide a more direct measurement but only for a single station over a rotated crop surface. Central facility daytime latent heat fluxes are lower than domain-averages for cases 1-3 by 57, 74 and 10 W m⁻², respectively (not shown). The lower values may arise from differences in the ECOR central facility fetch of surface properties compared to the EBBR network or other factors. Daytime means of Central Facility sensible heat fluxes differ from domain-averages (ARM_D-SF minus ARM_D) for cases 1-3 by 23, 36, and -49 W m⁻², respectively.

The absence of large-scale forcing (ARM_D-noLS) yields mostly flat simulations for all cases in terms of $\overline{q_v}$ and \overline{T} (Fig. 5). The $\overline{q_v}$ difference between ARM_D-noLS (blue line) and ARM_D (orange) clearly demonstrates the strong impact that horizontal advection flux divergence has on all three cases, ranging from least impact (but still significant) for case 1 to maximal impact for case 3. The timing of the Z_{LCL} diurnal cycles for ARM_D-noLS and ARM_D are similar in all cases because the diurnally varying surface fluxes are the same; timing differences are greatest during the air mass transitions in case 3. Comparing rows 2 and 3 in Figure 6 (ARM_D and ARM_D-noLS, respectively) illustrates some striking differences in cloud structure in the absence of large-scale forcings. Although case 1 cloud structures are little

affected (similar to *Brown et al.* [2002]), cloud structures are increasingly different in case 3 but they maintain similar tendencies, and clouds in case 2 are completely different—being primarily cumuliform where in the baseline they are stratiform (day 2) or absent (day 3). These results indicate the importance of large-scale forcings associated with changing synoptic states in cases 2 and 3.

When the surface fluxes in ARM_D are replaced with ECOR measurements from the central facility, the ARM_D-SF simulations roughly parallel the ARM_D time series for all variables in Figure 5 (red and orange lines, respectively). This indicates the relative consistency in the surface fluxes at the SGP compared to the domain average, as well as the importance of large-scale forcings in setting the overall trends in cases 2 and 3. The greatest differences are for RH and Z_{LCL} for cases 2 and 3 (Figs. 5g,h,k,l). For case 2, the ARM_D-SF RH is lower than ARM_D by up to 40% with a Z_{LCL} greater by 0.5-1.5 km, which correlates with respective differences (ARM_D-SF minus ARM_D) in latent and sensible heat flux, -74 and 36 W m^{-2} , respectively. The patterns for case 3 are opposite, with ARM_D-SF RH greater by 20% and Z_{LCL} less by 0.5-1.0 km indicating, in this case, that the shallower boundary layer, associated with the sensible heat flux difference of -49 W m^{-2} , dominates the minor reduction in moisture from the latent heat flux difference of -10 W m^{-2} to yield slightly greater RH. The ARM_D-SF cloud fraction differences parallel those in RH, with the case 2 day 2 cloud being much less (Fig. 6 f vs. h) and case 3 day 2 and 3 cloud being much more (Fig. 6 j vs. l). This suggests the potential value of including surface flux changes in ensemble forcings to account for temporal and spatial variability and measurement uncertainty, at least when the objective to constrain the simulations with observations.

7.3. Sensitivity to ensemble large-scale forcings

The differing impacts of the ARM, ECMWF, and MS-DA forcings on the simulated time evolution are considered next. We shift the baseline from the 300-km standard domain used in the last section to the 150-km reduced domain, which is more consistent with the GCM grid size used for the long-term climate simulations, such as those in the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [IPCC AR5, 2013]. Because the ARM reduced-domain forcings (ARM_d) were specially generated for this study, our analysis will continue to include results from the standard-domain ARM forcings (ARM_D) that are routinely available as a reference. DHARMA is used for all simulations except MS-DA, for which the WRF model is used. The WRF model, WRF-FASTER as an LES as described in *Endo_PartII*, uses the same ‘trial’ domain size and grid resolution as DHARMA and only warm cloud physics. When tested using the same forcing data, WRF closely tracked DHARMA results, indicating that the models serve equally well to test case specifications (not shown; see also part II).

7.3.1. General tendencies

The ensemble simulations capture the observed features with varying success. Results are presented in terms of boundary layer atmospheric state variables (Fig. 7), time-height cloud masks (Fig. 8), and cloud fraction and $\overline{\text{LWP}}$ (Fig. 9). Compared to properties that are highly variable, gradually varying properties are better represented, such as the case 2 $\overline{q_v}$ (Fig. 7e) and all \overline{T} s (Fig. 7b,f,i), except for the evaporative cooling dips seen in case 2 from 27.3-27.5 May and in case 3 at 6.5 May. However, the higher variability in other properties appears commonly underestimated throughout the forcing ensemble.

In case 1, for example, the observed diurnal variability in $\overline{q_v}$ and $\overline{\text{RH}}$ (Fig. 7a,c) is best captured by MS-DA (green), followed by ECMWF (blue), and is underestimated by ARM_D and ARM_d (orange and red). A similar pattern is seen in the Z_{LCL} diurnal cycle (Fig. 7d) where

all forcings capture the peaks in Z_{LCL} , but only MS-DA is able to represent the nighttime minima. Since all simulations use the surface forcing from ARM_D, the ability of MS-DA to represent the deep stabilization at night is likely due to its representation of near-surface moisture advection. This suggests that assimilating ARM conventional observations and satellite radiance might be helpful in capturing finer scale flows to create realistic large-scale forcing for this case. The variational analysis used in ARM represents a column balance, not specific terms as in data assimilation, so perhaps it is not surprising that higher frequency spatiotemporal variations are not well captured. Fine-scale assimilation appears most advantageous for the weakly forced systems in case 1, but is less dramatically so in case 2 (Fig. 7h) that was previously discussed as being advection driven. (The variability in case 3 will be discussed later.) Interestingly, the reduced-domain (150 km) ARM_d does not improve simulation results over the standard-domain (300 km) ARM_D and, in fact, can even degrade the simulation in terms of these metrics. This might indicate that the current column constraints are best when applied/averaged over larger domains. Improvement of the variational analysis to three-dimensions is a topic of ongoing research. However, representing this atmospheric variability is only part of the picture and, as will be seen later, cloud onset and cloud properties can be captured comparably by the two forcings.

7.3.2. Effects of initial condition specifications

Simulations are essentially identical for the first 6-12 h of day 1 for each case, regardless of forcing (c.f., Fig 7, row 1). This stems from the simulations being initialized by the same SGP central facility sounding and the time required for the effects of differences in the large-scale forcings to accumulate. Thus, initial conditions seem to be more important to short-term simulations than the source of forcing (see also ARM_D-noLS versus ARM_D in Fig. 5).

Simulations diverge earliest in case 3 (after ~6 h in Fig. 7i), indicating the greatest differences among the forcings for that case. Once divergence occurs, the model states maintain independent identities, even when the simulations are relaxed to the same state profile per case (see section 5.4). We note that, while it might be more consistent with our use of the ARM_D surface forcing to initialize with the ARM_D profiles instead of the sounding, we found that simulations performed better using the sounding, presumably because details in the vertical structure are smoothed out with the horizontal averaging of the RUC profiles across the 300 km domain.

Given the importance of the initial sounding, uncertainties in its representation may have important short-term consequences as well. For example in case 1, \bar{T} are biased low by 3 K compared to the Raman data in all simulations before 23.0 May (Fig. 7b). Since simulated \bar{q}_v initially agrees well (Fig. 7a), this leads to a consistently positive $\overline{\text{RH}}$ bias (Fig. 7c) for the first 4-6 h. A closer examination of the sounding and Raman lidar observations (not shown) finds close agreement of the boundary layer properties at the time of the launch, but that the temperature quickly changes by one hour after launch to the 3 K difference noted above. This suggests the need for ensembles of initial condition profiles for such boundary layer simulations to capture the variance around the time of the launch, which might be caused by local-scale differences across the region.

The issue of initial conditions also applies if clouds are present at the beginning of a simulation. For example in case 1, early morning cloud is evident about 7 LST (Fig. 8a) with cloud fractions of 0.70-0.95 and $\overline{\text{LWP}}$ peaking at 400 g m^{-2} (Fig. 9a-d). However, the 11:30 May sounding used for initialization is subsaturated in the boundary layer (Fig. 10) consistent with the sub-cloud $\overline{\text{RH}}$ of 60-70% seen in Figure 7c. Consequently, cloud is absent for the first 3-4 h of

simulation; afterwards, \overline{LWP} is uniformly underestimated by 100 g m^{-2} for the next 3 h, and cloud fraction by 0.5 for day 1 (Fig. 9a-d). Note that such environmental conditions are not uncommon at the SGP, since elevated layers of residual moisture from the nocturnal boundary layer may remain in the early morning. Since these moisture layers are above the growing unstable boundary layer, they can develop cloud in advance of the main cumulus field of interest to us here. Case 3 shows a different example where a significant stratiform cloud is present at 6.5 May (Fig. 8k), with cloud fraction of 0.8-1.0 and \overline{LWP} of 200 g m^{-2} (Fig. 9i-l). However, contrary to the Raman moisture data (Fig. 7k), the initial sounding is unsaturated (Fig. 10) and simulations take about 3 h to spin the cloud up to the point of matching the observed cloud fraction and \overline{LWP} . Thus, properly producing the morning and mid-day cloud would have required iterative ad-hoc adjustments to the observed sounding to reflect the variability of the observed cloud field shortly before and after the sounding and its effect on the RH profile. This could involve increasing the RH profile in the initial conditions, consistent with the Raman lidar, and/or adding cloud water using the ARSCL profile and LWP measurements.

7.3.3. Relaxation impacts

Applying 12-h thermodynamic relaxation in the simulations leads to minor changes in the boundary layer atmospheric properties, as seen by the small differences between the solid and dashed lines in Figure 7. This may be because of a small amount of drift for the relaxation to act upon, the relatively long 12-h thermodynamic relaxation timescale used, and/or because relaxation is only fully applied from 600 m upwards (linearly increased from 400 m) while the boundary layer properties are calculated for a layer between 300-500m. However, relaxation can play a significant role in simulated cloud properties, particularly by reducing errant cloud occurrence. This is seen in the 2d cloud masks in Figure 8; green indicates where the

simulations with and without relaxation both have cloud, red is where cloud is only present without relaxation, and blue is where cloud is only present with relaxation. Thus for case 1 (Figs. 8b-e), the difference between the green and red areas shows that the cloud vertical extent is often dramatically reduced when relaxation is used which increases consistency with observations, particularly for MS-DA. For case 2 (Figs. 8h,i), relaxation suppresses the errant high-level cloud in ARM_d and ECMWF. In case 3 (Figs. 8l-o), relaxation tends to reduce cloud prior to the cold front passage at 8.5 May. Note that the ARM_D cases overall have the least differences between the relaxation experiments than do ECMWF or MS-DA. This may be because, when mixing ARM boundary conditions with ECMWF or MS-DA flux divergences, more drift may be expected since the forcing is not entirely self-consistent. Overall, relaxation has its greatest effect in the last two days of simulation and little effect within the first 12 h. This pattern is also seen in Figure 9 where cloud fraction without relaxation (row 1) and with relaxation (row 2) are essentially the same for the first 12 h of simulation, as is true for $\overline{\text{LWP}}$ without relaxation (row 3) and with relaxation (row 4). That it takes more than 12 h for relaxation to have an impact is indicative of the 12-h relaxation time used; however, as discussed in the previous section, it also takes roughly that long for small errors in the advective forcings to accumulate to the point where relaxation can act to offset the resulting drift.

Although relaxation can improve cloud simulations, its overall impact can be mixed. For example, relaxation reduces the case 1 noontime $\overline{\text{LWP}}$ at 23.8 May by 50 g m^{-2} , bringing it in close agreement with the MIXCRA retrievals (Fig. 9c vs. d). However, cloud fraction without relaxation was already in good agreement with observations and, with relaxation, became biased low by at least 0.4 (Fig. 9a vs. b). Since these relaxation effects are similar for all forcings, it suggests that the common cause could be the ARM_D profile to which all simulations are

relaxed. Generally speaking, relaxation is only as effective as the quality of the forcing to which it is relaxed.

For case 2, the relaxation has little effect on the low-level clouds as evidenced by the green areas (Fig. 8g-j). In all simulations, a single-layered nighttime stratus forms by the end of day 1 (27.0-27.2 May) that later develops into a two-layered stratus cloud by early morning on day 2 (27.5-27.7 May). Figure 8f shows that the lower cloud layer (< 2 km) is present in observations but that the upper layer is not. Interestingly, relaxation hardly affects this feature in the cloud mask, or in cloud fraction and \overline{LWP} (Figs. 9e-h). The simulated \overline{LWP} matches observations during nighttime from 27.0-27.5 May (Fig. 9 g,h), indicating the issue is in cloud location rather than \overline{LWP} amount. The double-layered feature is strikingly robust, appearing to some extent for all forcings and regardless of whether or not relaxation is used. If the upper cloud were too thin, it would not be detected by the cloud radar; however, simulated liquid-water content has maximum values towards the top of each layer (not shown), suggesting that both layers are substantial enough to be detected. Although the common cause of the upper cloud is puzzling, relaxation would not affect it if it also exists in the ARM_D profile used for relaxation. The development of each individual cloud feature is beyond the scope of this study, but we speculate that this feature might result from longwave cooling at the top of the inversion creating cloud-top instability.

7.3.4. Atmospheric horizontal gradients

The air mass changes in case 3 present atmospheric gradients that the forcings must capture to simulate cloud features similar to observed. As seen in Figure 7i, all forcings misrepresent the $\overline{q_v}$ variability. ECMWF and MS-DA are able to capture the $\overline{q_v}$ dip to 7 g kg^{-1} between 7.0 and 7.5 May, but grossly underestimate the following increase to 20 g kg^{-1} , reaching a plateau at 13-

14 g kg⁻¹. ARM forcings capture the magnitude of the dip from 7.0-7.5 May, but not the sharpness, and the forcings do capture the increase to 20 g kg⁻¹, but it lags observations by 12-18 h. Z_{LCL} is poorly represented by all forcings after 6.9-7.1 May, likely indicating issues representing the near-surface air flow. The stratus cloud onset at 7.5 May is delayed by 3-4 hours by the simulations (Figs. 8k-o), and its cloud fraction (Figs. 9i,j) and \overline{LWP} (Figs. 9k,l) are underestimated by 0.4-0.8 and at least 100 g m⁻², consistent with the dry bias in the moisture fields. Similar cloud fraction biases occur 8.0-8.5 May, but \overline{LWP} agreement for some forcings is reached by ~8.3 May.

It is notable that all three state-of-the-art forcings apparently fail to capture the conditions of case 3 beyond the first 12 h. This is discussed further in section 8.

7.4. Ensemble simulations for the flight periods

An objective of this work is to identify case specifications that best represent the atmospheric and cloud macroscopic observations during flight periods for use in more intensive study with in-situ cloud microphysical observations. As has been seen, many factors affect the time-evolution of the simulated properties. Here we use averages of the time series over the flight periods to provide a succinct assessment of specification quality. We examine the relationship between the flight-period-averaged cloud fraction and \overline{LWP} , the two most basic cloud macrophysical properties for which observations are available.

The flight-time-averaged cloud fraction and \overline{LWP} are presented as scatter plots in Figure 11, where each point is an average over the flight period for a given simulation. Results for the 5/8 flight are not shown because most simulations are invalid following the passage of a squall line to the north of the central facility at 8.5 May. (Observations for that flight are $\overline{LWP}=20.3 \pm 2.5$ and the cloud fraction range is 0.81-0.93.) As discussed earlier, day 1 for case 1 (5/22) contains

a residual layer of moisture that activates in the early morning in advance of the main cumulus field. This cloud is not representative of the cumuli of interest here that are coupled to the surfaced processes and, as seen in *Endo_PartII*, exhibit a different cloud microphysics character from the rest of the field. An attempt is made to approximate the main cumulus cloud field properties by removing the largest contribution of the residual-layer cloud. This is done by limiting the ARSCL height field to lower levels, and restricting the TSI cloud fraction and LWP observations to later times when the residual cloud loses distinction from the rest of the field. That approximation is represented by the vertical gray bar in Figure 11a, where the long vertical extent of the bar reflects the more limited ability to restrict $\overline{\text{LWP}}$ than cloud fraction.

Many of the features previously discussed are seen in these plots. For the case 1 cumulus (Figs. 11a-c), all simulations underestimate cloud fraction on day 1 (5/22) compared to observations for the full flight period, but cluster around the approximation that removes residual-layer cloud. Simulations without relaxation (boxes) agree well with observations on day 1, but increasingly drift from them on successive days. Relaxation (pluses) helps maintain all simulations closer to observations, although cloud fraction is underestimated by > 0.1 on days 2 and 3. For the case 2 cumulus and drizzling stratus (Figs. 11d,e), the general overestimation of $\overline{\text{LWP}}$ discussed earlier is seen, but cloud fraction is well represented by all simulations except for day 2 of ARM_D-SF. For this case, the MS-DA_d runs with and without relaxation agree best with observations. For the case 3 variable cloud types (Figs. 11f,g), all simulations but ECMWF perform well on day 1, and none of the simulations do well with the atmospheric gradients on day 2. Although ECMWF is an outlier here and on day 2 of case 2, it generally does remarkably well given that its forcing only benefits from using ARM soundings and satellite data but not the other local ARM observations that help constrain ARM and MS-DA. Interestingly, although

relaxation helps prevent drift in the weakly forced state in case 1, it does not provide a marked improvement over the unrelaxed simulations for the other cases.

8. Summary and Conclusions

The goal of this study is to provide a diverse set of observation-constrained cases of continental cloudy boundary layer evolution for use by the modeling community to improve understanding of boundary layer clouds, aerosol influences upon them, and their representation in cloud-scale and global-scale models. Developing cases such as these is intended to help address a general lack of detailed continental shallow cloud cases within the community. The multi-day case study periods span the temporal evolution of cumulus, stratiform, and drizzling boundary layer cloud systems, representing mixtures of cloud types and transitions that are commonly observed. The cases share a standardized modular specification for ease of use by LES and SCM modelers. We note that setting up case studies for use in LES and SCMs requires some upfront thought to the configuration. For example, while shallow convection large-eddy simulations might use a model top of only ~ 5 km for computational efficiency (*e.g.*, *Endo_PartII*), SCMs require the whole atmospheric profile and it would be difficult to apply a patch above 5 km after the fact. The case study data will be available from the ARM archive as a Principal Investigator (PI) Data Product that is freely available to the public. Further, to make the setup of cases more flexible for being restarted at 24 and 48 hours from time zero for each case, additional soundings and clear-sky longwave radiative transfer calculations are provided (the already provided large-scale forcing profiles and surface fluxes are applicable continuously). Finally, the presented simulation results at 10-min resolution are available upon request from the authors.

In addition to the extensive SGP routine observations, the cases benefit from aircraft in-situ constraints. A distinctive aspect of these specifications is the detailed, observation-based aerosol characterization. Measurements of aerosol number size distributions are fit to lognormal distributions for concise representation in models. Aerosol hygroscopicity, κ , is derived from CCN data, which yields values ~ 0.10 —lower than the 0.3 typical over continents and suggestive of a large aerosol organic fraction. The aerosol specification is used in part II of this series [Endo_PartII] to examine details of the simulated cloud microphysics in comparison with in-situ measurements in the case 1 cumulus . In part III [Lin_PartIII], the large-eddy simulations and aircraft-measured dynamical properties are used to examine the shallow cumulus parameterization used in the CAM5 SCM. The companion studies demonstrate the utility of the case study modular specification, and the three-paper series collectively provides an example for case 1 of the intended end-to-end analysis of continental boundary layer cloud multi-scale processes, from observations to LES to SCM scales.

Another distinctive aspect of this study is the use of an ensemble set of large-scale forcings from three distinct methodologies, derived from ARM continuous forcing data, ECMWF forecasts, and a multi-scale data assimilation system. The considerable computational load to run LES on the ensemble members and their variations (relaxation and sensitivity studies) is made tractable (although still considerable) by using ‘trial’ LES, where run times that are one-fifth of the full-sized LES are achieved through modest reductions in grid resolution and domain size but still capture the bulk behavior of the simulation. The trial LES, evaluated using measurements of boundary layer thermodynamics and cloud macrophysical properties, provide an efficient means of determining common factors affecting simulations and when results justify full-scale LES runs for more detailed analysis.

Although the boundary layer is strongly forced by surface fluxes over the diurnal cycle, horizontal advection flux divergence is shown to have a range of impacts, from minimal for the locally forced cumulus in case 1 to maximal for the air mass transitions in case 3. Trial simulations can help determine when large-scale forcing can be ignored (i.e., the difference between the ARM_D and ARM_D-noLS runs). We note that all three forcings apparently fail to reproduce the conditions of case 3 beyond the first 12 h, particularly the sharp $\overline{q_v}$ increase after 7.5 May. It is possible that grids used by regional-scale models to generate forcings are too coarse to represent the tight gradients or high-resolution transients of importance here, or that details of location and timing are sufficiently chaotic that they were not reproduced in a particular model. While the 12-km RUC grid used by ARM and the 22-km ECMWF grid are well-resolved by global model standards, they are still sufficiently coarse to require parameterized cloud processes. However, even the MS-DA, which uses a 2-km resolution in its inner domain, also did not represent the sharp gradients observed. This shortcoming has important ramifications to the generation of forcings needed by cloud modeling studies to improve process understanding and physical representations in climate models, at least when the objective is to use observations to constrain model performance. For now, pending further investigation, it seems advisable to avoid cases with large gradients that sources of large-scale forcings cannot capture.

Simulation quality depends on different factors at different times in the simulations. Initial conditions dominate over the large-scale forcing < 6 h, and differences among the forcings becomes important after 6-12 h. Thermodynamic relaxation can circumvent some uncertainty in the forcing datasets to improve cloud simulations; however, it is not a blanket remedy, as the relaxation term does not represent a real physical process [Randall and Cripe, 1999] (see

discussion in section 5.3) and simulation improvements can be mixed. Generally speaking, relaxation can act only after errors in the advective forcings have accumulated sufficiently to be offset, and the level of improvement is contingent on the quality of the profile to which the simulation is being relaxed. Relaxation helps prevent the drift in the weakly forced state in case 1, but does not provide a marked improvement for the other two more strongly forced cases.

Ultimately, since these forcings, which are state-of-the-art, fall short for the purpose of driving boundary layer cloud simulations with gross thermodynamic properties as observed, further study would be required to establish a methodology consistently capable of doing so. While this work surveyed the effects of initial conditions and different large-scale forcings on the simulations, a careful, systematic modification of different factors is required (such as that used in section 7.2) to determine a methodology for obtaining the best forcing to match the timing of cloudy periods and their properties. Particularly for weakly forced systems, the effects of fine-scale subgrid effects on the domain seem to be important to represent. Recall that surface fluxes are imposed from the forcing data and therefore prevent the connection between clouds and surface fluxes. An important factor to capturing subgrid effects likely involves a coupled surface-atmosphere treatment. Another input may be the calculation of forcings required to reproduce a particular observed condition.

Given that initial conditions dominate the beginning of the simulation over other factors investigated here, special consideration should be given to their bias error and representativeness of the variability within the domain. For example, recall that case 3 has a significant stratiform cloud but the initial sounding is unsaturated. At the SGP site, the sounding below cloud base can be verified using Raman lidar temperature and water vapor, and those measurements may also provide information to optimize a correction when needed. Should cloud be present, ARSCL

observations can indicate the atmospheric layers that must be saturated. However, some testing of this idea found that saturating the layer had little effect, suggesting that cloud water too must be added, which introduces additional methodological complexity. For example, an approximate correction to the initial conditions could be formulated based on the ARSCL cloud mask, LWP measurements, and the assumption of an adiabatic profile. Should relaxation be used, the same procedure could be applied to soundings later in the simulation since relaxation is only as effective as the quality of the profile to which it is relaxed.

Representing the variability in initial conditions caused by local-scale differences across the region is more challenging. An obvious starting point is to use ensembles of the profiles and surface forcings. For example, properly producing morning and mid-day cloud may require adjustments to initial profiles to reflect the variability of the cloud field shortly before and after a sounding and its effects on the RH profile. The horizontal variability of atmospheric conditions can be characterized using Raman lidar profiles around the time of the sonde (*e.g.*, ± 1 h). Similarly, the variance in the surface flux measurements within the domain should be considered since surface fluxes can make a big difference and vary substantially over the domain around the central facility. However, blindly running ensembles for all possible permutations would be computationally prohibitive and, further, it would ignore the coherency within the observations vertically and horizontally. While each atmospheric profile should be maintained as an entity, it should also be linked to the surface fluxes. Directly observing all of the profile and vertical flux pairings is beyond the capabilities of even a site as heavily instrumented as the SGP. However, sufficient sonde, Raman lidar and surface flux measurements probably do exist to attempt a statistical (Bayesian) mosaic of the structures.

Appendix A: RACORO Observations

The CIRPAS Twin Otter flew a comprehensive payload of instruments listed in Table A1. For details on instrument status, data recommendations, and flight plans see the RACORO data guide [Vogelmann, 2012]. To accurately measure quick-varying cloudy boundary layer properties with instrumentation robust enough for the operational constraints of an extended-term field campaign, when possible, a pair of instruments was deployed—a slow-response measurement providing the needed accuracy and a fast-response measurement quantifying the variability (see discussion in Vogelmann *et al.* [2012]). Highlights of the measurement payload are as follows (see Table A1 for instrument acronyms).

- Cloud microphysics observations include multiple measures of drop size distribution (CAS, FSSP, 1D-CIP, 2DS, 2D-CIP), and of LWC (from Gerber and SEA LWC bulk measurements, and integration of CAS [preferred] or FSSP size distributions).
- Aerosol physical measurements include: CPC total number concentrations (for diameter > 3, 10 and 15 nm), aerosol size distributions from a PCASP (100–2,200 nm) and an SPMS (12–600 nm), and CCN concentrations at five supersaturations (0.2, 0.28, 0.4, 0.57 and 0.8%).
- Radiative observations include: cloud extinction measured by a CIN, broadband solar irradiances (a modified CM22 and a fast-response SPN1), thermal irradiances (a modified CG4), five narrowband spectral irradiances including a 1.6 μm channel (MFR), infrared radiance (IRT), and high-resolution spectral irradiances and radiances (HydroRad-3).
- Atmospheric state observations include turbulence from a gust probe, temperature (Rosemount and Vaisala), and water vapor concentration (EdgeTech and CR2 Chilled-Mirror Hygrometers) including an ultra-fast measurement by a Diode Laser Hygrometer at 100 Hz.

Appendix B: Model Evaluation Data

The surface-based observations of bulk cloud and boundary layer properties used to evaluate the simulations and their uncertainties are described here.

- a. Liquid-water path. Two cloud LWP retrievals are used. MWRRet [Turner *et al.*, 2007a] uses surface-based microwave brightness temperatures (23.8 and 31.4 GHz) that have been bias corrected, yielding an LWP uncertainty of 20-30 g m⁻². MWRRet is routinely available and is used for analyses that cover the full study periods. Much more accurate LWP retrievals are available for limited periods from the Mixed-phase Cloud Retrieval Algorithm (MIXCRA) [Turner, 2007]. It uses the microwave brightness temperatures and surface-based infrared radiances (8–13 μm and 3–4 μm) that, combined, yield superior retrieval to those using infrared or microwave radiances alone, in terms of uncertainty and LWP range. This is by virtue that infrared radiance provides sensitivity at small LWP (< ~60 g m⁻²) where the microwave retrievals have large uncertainty, and microwave radiance is sensitive to a wide range of LWP (5 to 1000 g m⁻²) that is far beyond the 60 g m⁻² where the infrared loses sensitivity. The MIXCRA algorithm reports one-sigma uncertainties that are less than 30% for LWP from 1 to 5 g m⁻², and are less than 10 to 20% for LWP greater than 5 g m⁻². Our analysis removes LWP less than 1 g m⁻² because they may result from haze layers (hydrated aerosol) or thin cirrus. For comparison to LES results, domain-averaged LWP, $\overline{\text{LWP}}$, is determined by averaging clear-sky and retrieved cloud LWP over a time period (*e.g.*, a flight period) and applying the frozen turbulence assumption (or Taylor hypothesis). For either of the retrievals, a leading source of uncertainty for broken clouds is sampling and applicability of the frozen turbulence assumption.

b. Cloud fraction. Cloud fraction estimates depend on the sensitivity and field of view of the sensor employed [e.g., Wu *et al.*, 2014] so two measurements are used to approximate the measurement uncertainty. The total sky imager (TSI) is a hemispheric-viewing camera providing retrievals of fractional sky cover during daytime for ‘opaque’ and ‘thin’ clouds. Fractional sky cover can be overestimated due to scattering from cloud edges, particularly from clouds on the horizon; this effect is minimized by only using measurements from the 100° field-of-view. Cloud fraction is derived as 10-min averages of the TSI opaque fractional value, which is most relevant to the boundary layer clouds of interest. (Our averaging period is slightly less than the optimal 15-min interval recommended by Kassianov *et al.* [2005] for slightly better temporal resolution.) The other cloud fraction estimate is derived from the ARM Active Remote Sensing of Clouds (ARSCL) product [Clothiaux *et al.*, 2000], which determines cloud layers from combined observations from a micropulse lidar and millimeter wavelength cloud radar as clouds advect through a narrow column above the instruments. Cloud fraction is derived from the cloud frequency per time interval as described in Xie *et al.* [2010], which assumes a horizontally uniform cloud field distribution (i.e., the frozen turbulence assumption). Cloud fractions are ten-minute averages of fractional occurrence computed from the vertically-resolved ARSCL cloud mask for clouds lower than 5.5 km. In simulations, ‘cloud’ is defined as $LWP > 1 \text{ g m}^{-2}$ to approximate the lower detection limit of the sensors (e.g., MIXCRA). Below this value, MIXCRA retrievals might measure haze or thin cirrus. The 1 g m^{-2} cutoff is somewhat arbitrary but yields cloud fractions similar to when 0.1 g m^{-2} is used. Note, however, that a 0 g m^{-2} cutoff can yield greater cloud fractions by up to 0.2.

- c. Lifting condensation level height. Lifting condensation level (LCL) height is determined from surface-air observations of relative humidity and temperature, as the altitude where the surface-air moisture equals saturation following a dry-adiabatic ascent. The domain-averaged LCL height, Z_{LCL} , and its uncertainty are computed from the mean and standard deviation of five ARM and Oklahoma Mesonet (<http://www.mesonet.org/>) stations closest to the SGP central facility. The same calculation is applied to the lowest air layer in the simulations, enabling a consistent observation-model comparison.
- d. Boundary layer moisture and temperature. A Raman lidar [Goldsmith *et al.*, 1998] provides high frequency vertical profiles of the boundary layer q_v [Wulfmeyer *et al.*, 2010], temperature [Newsom *et al.*, 2013], and relative humidity (RH) computed from these measurements. The native temporal and vertical resolutions of the measurements are 10-s and 7.5-m, which are averaged and provided to the community at 10-min and ~ 75 m resolution. The automated processing algorithms [Turner *et al.*, 2002; Newsom *et al.*, 2013] provide one-sigma uncertainties for the random errors in q_v and temperature of about 0.2 g kg^{-1} and 3.5 K. We assign the q_v uncertainty to be its random error or 5% of the value, whichever is larger, since there is some calibration uncertainty not included in the random error. RH uncertainty is propagated from its component errors. Measurements are valid only below cloud base, above which the lidar signal saturates. Tenuous cloud below the point of saturation can bias q_v measurements by contributing liquid-water Raman scattering to the water vapor signal [Melfi *et al.*, 1997]. Range gates affected by cloud or below-cloud effects are screened by removing the measurements at cloud-base height plus one range gate lower, and removing range gates where the q_v random error is greater than 1.5 g kg^{-1} plus one range gate lower. This screening does not always remove the effects of liquid water (drizzle),

which is detected by when RH greatly exceeding 100%. In these cases, the lower uncertainty bound for q_v is adjusted downwards to where 100% RH would be reached assuming that the mean temperature is accurate (temperature channels are not affected by liquid water Raman scattering). Post-screened data 300 to 500 m above the surface are averaged using a one-hour moving window to produce mid-boundary-layer-averaged $\overline{q_v}$, \overline{T} , and \overline{RH} , where eight valid measurements are required per average. The 500-m top altitude is chosen because it is generally being below cloud base for our three cases. The 300-m floor is the minimum altitude for valid Raman lidar retrievals because of receiver overlap considerations. The high spatial and temporal resolution of the Raman lidar data are valuable for assessing the simulated boundary layer evolution; however, it is a point measurement that cannot represent the total variation across the model-simulated domain.

Appendix C: Measurements of Aerosol Size Distribution and Cloud Condensation Nuclei

Aerosol size distributions were measured by the CIRPAS Twin Otter using a Scanning Mobility Particle Sizer (SMPS) and a Passive Cavity Aerosol Spectrometer Probe (PCASP)-100X with a SPP-200 electronic upgrade. The SMPS makes size-resolved measurements of dry aerosol particle diameter (at 10% relative humidity) from 12-600 nm every ~60 s, and the PCASP from 100-2,200 nm at 1 Hz. The PCASP was calibrated using polystyrene latex (PSL) beads that have a real refractive index of 1.59 and yields number concentrations in good agreement with the SMPS in the size overlap region of the two instruments (100-600 nm).

Aerosol size distribution profiles were generated from measurements made during aircraft spirals performed at the beginning and end of each flight. Each spiral took about 600 s and ranged from 150 m above-ground level (AGL) to about 450 m above cloud top [Vogelmann, 2012]. SMPS and PCASP data were binned into 100-m height intervals and cloud screened using LWC measured at 1 Hz by the CAS. A SMPS or PCASP sample was removed if LWC $\geq 0.001 \text{ g m}^{-3}$ during its measurement (i.e., at anytime during the 60-s SMPS cycle, or for the concurrent 1 Hz PCASP measurement). An average size distribution was first computed separately for the SMPS and PCASP data for each 100-m interval. Then, a single size distribution is generated by combining the SMPS and PCASP data via a weighted average of the number concentrations in the size overlap region, where the weights transition from 1.0 SMPS and 0.0 PCASP at 100 nm to 0.0 SMPS and 1.0 PCASP by 600 nm.

CCN measurements used in this analysis were made at multiple supersaturations (SS_i) using a Droplet Measurement Technologies (DMT) Dual-Column CCN Spectrometer (CCN-200). One column was held constant at 0.2% supersaturation and reported at 1 Hz. The second column scanned downwards through five supersaturations (0.80, 0.57, 0.40, 0.28, 0.20) in about 25 min

before initiating a new cycle. The scanned results are also provided at 1 Hz, but measurements are only valid when instrument stability criteria are met (such as the temperature difference across the column plates yielding the desired supersaturation). The nominal values of supersaturation were estimated from instrument settings and calibration information; however, when the instrument operates under conditions different from which it was calibrated (*e.g.*, different pressures), the actual supersaturations can be significantly different and vary with environmental conditions. Actual supersaturations were calculated for the two columns using a detailed model of the instrument [*Lance et al.*, 2006] by E. Andrews and J. Ogren of NOAA (available in the RACORO IOP Data Archive as ‘Adjusted Cloud Condensation Nuclei Concentrations’). For example, the five scanning supersaturations from the 22 May flight are decreased from their nominal values of (0.80, 0.57, 0.40, 0.28, 0.20) to their actual values of (0.53, 0.40, 0.30, 0.23, 0.19), where the greatest differences occur at the highest supersaturations.

Screening criteria were applied to the CCN to remove spurious points. Similar to the cloud screening for the PCASP and SMPS, CCN measurements are removed when $LWC \geq 0.001 \text{ g m}^{-3}$. (This is conservative as the CCN measurements did not appear to be affected by cloud droplet ‘splashing’ [*E. Andrews*, personal communication].) To assure stability for the CCN measurements made by the scanning column, we verified that the temperature across the column plates was correct for the given supersaturation (*i.e.*, not in transition between temperatures). Also, after a given supersaturation was reached, the first 1-min CCN average was removed because it often was not yet stable. This procedure was verified through comparison of the 0.2% supersaturation CCN values from the scanning column to those from the column fixed at 0.2%.

Appendix D: κ Derivation from Aircraft Observations

Based on κ -Köhler theory [Petters and Kreidenweis, 2007], κ relates the minimum saturation ratio, $S_{c,i}$, required to activate a dry aerosol particle diameter, $D_{c,i}$,

$$D_{c,i} = \left(\frac{B}{\kappa (\ln S_{c,i})^2} \right)^{1/3}, \quad (\text{B1})$$

where i is the index for saturation ratio. The constant B is,

$$B = 4 \left(\frac{4\sigma_{s/a} M_w}{3RT\rho_w} \right)^3 10^{18},$$

where $\sigma_{s/a}$ is the surface tension of the solution droplet, M_w and ρ_w are, respectively, the molar mass and density of water, R is the universal gas constant, and T is the absolute temperature. Following Petters and Kreidenweis [2007], $\sigma_{s/a}=0.072 \text{ J m}^{-2}$ and $T=298.15 \text{ K}$, which yields $B=1.355 \cdot 10^{-9} \text{ } \mu\text{m}^3$.

CCN were measured at multiple supersaturations (SS_i) using a DMT Dual-Column CCN Spectrometer, where one column was held constant at 0.2% supersaturation and the second column scanned downwards through five supersaturation steps (0.53, 0.40, 0.30, 0.23, 0.19) in about 25 min. Details on the CCN instrumentation and data screening are provided in Appendix C. Valid 1 Hz measurements were averaged into 60-s intervals.

Assuming that all particles are an internal mixture of the same chemical composition and therefore κ value, at a given critical saturation ratio $S_{c,i}$ (or critical supersaturation, $S_{c,i}-1$), all particles with diameter greater than $D_{c,i}$ (Equation [B1]) serve as CCN. The CCN concentration at $S_{c,i}$ is derived by integrating, from $D_{c,i}$, the fitted size distribution (equation [1]):

1168

$$N(D_{c,i}) = \int_{D_{c,i}}^{\infty} \frac{dN}{d \log_{10} D} d \log_{10} D = CCN_i. \quad (B2)$$

1169

1170

1171

1172

1173

1174

1175

1176

1177

A κ is determined for each spiral profile from the measured pairs of CCN_i and SS_i , and the 3-mode fits of the aerosol size distributions from section 3.1 with fixed geometric mean diameter and standard deviation. Since CCN_i measurements per SS_i were only made once every 25 min, each flight provides a sample of the vertical variation of CCN_i , but all SS_i are not necessarily measured at the same height intervals. Thus, we combine the CCN_i per SS_i into a single flight average and, correspondingly, perform our analysis using the height-average of the aerosol size distribution profile. Since there are two spirals per flight each yielding a profile of size distribution, κ is derived for each spiral using the same flight-averaged CCN_i , which is found to yield essentially the same κ value (discussed in main text).

1178

1179

1180

The lower limit of the integral in equation (B1), $D_{c,i}$, is a function of κ (see equation [B2]). Thus, κ is solved using a Levenberg-Marquardt least-squares fit [Markwardt, 2009] of the equation,

1181

$$\begin{bmatrix} N(D_{c,1}) \\ N(D_{c,2}) \\ \dots \\ N(D_{c,5}) \end{bmatrix} = \begin{bmatrix} CCN_1 \\ CCN_2 \\ \dots \\ CCN_5 \end{bmatrix}.$$

1182

1183

As discussed in the main text, the observed CCN spectra agree well with that computed from the resulting κ and the height-averaged size distributions, supporting the analysis procedure.

1184

1185

1186

A second analysis checked this result by determining κ directly from the observations, without the use of the fitted size distributions or least-squares fitting. $D_{c,i}$ was determined from the SMPS data by integrating the size spectrum from large to small sizes until the number of

1187 particles equaled CCN_i ; the smallest SMPS diameter in the integration was taken to be $D_{c,i}$. For
1188 each SS_i , κ were computed using the values obtained for $D_{c,i}$ and $S_{c,i}$ in equation (B1). The
1189 results yielded similarly low κ , although with a bit more variability (not shown).
1190

Acknowledgements

Data used in this article are from the U.S. Department of Energy SGP ARM Climate Research Facility (available from <http://www.archive.arm.gov>) and the AAF RACORO Campaign (available from <http://www.arm.gov/campaigns/aaf2009racoro#data>). We thank the entire RACORO team: the RACORO scientific steering committee, Haf Jonsson for the analysis and processing of the Twin Otter flight data and recalibration of the PCASP, the instrument mentors for their analysis and processing of data, and the DOE ARM Aerial Facility for its coordination of RACORO. We also especially acknowledge Don Collins for guidance using the SPMS data, David Cook provided informative discussions regarding the surface roughness length over the SGP, Krista Gaustad and Laura Riihimaki for special processing of MWRRet data for 8 May, and the Raman lidar mentor team of Chris Martin, John Goldsmith, and Rob Newsom for their efforts in maintaining the Raman lidar. Ozone measurements from the Ozone Monitoring Instrument (OMI) were provided by the NASA/GSFC TOMS Ozone Processing Team (OPT) and obtained via the ARM External Data Center. We would like to thank three anonymous reviewers for their thoughtful comments on the manuscript. This research was supported by the U.S. Department of Energy Science Office of Biological and Environmental Research Program under the following grants/contracts: the Earth System Modeling Program via the FASTER Project (AMV, TT, WL, SE, YL, SF, ZL, MZ, MK), and the Atmospheric System Research Program via DE-SC00112704 (AMV, YL, JW), DE-SC0006988 (AMF and ASA), and DE-SC0006898 (DDT). AMF and ASA used resources of the National Energy Research Scientific Computing Center, which is supported by the Office of Science of the U.S. Department of Energy under Contract DE-AC02-05CH11231, and the NASA High-End Computing (HEC) Program through the NASA Advanced Supercomputing (NAS) Division at

1214 Ames Research Center, and received additional support from the NASA Radiation Sciences
1215 Program. Work at LLNL was supported by the DOE ARM program and performed under the
1216 auspices of the U. S. Department of Energy by Lawrence Livermore National Laboratory under
1217 contract No. DE-AC52-07NA27344 (YZ, SX).

1218

REFERENCES

- 1219
- 1220 Ackerman, A. S., O. B. Toon, and P. V. Hobbs (1995), A model for particle microphysics,
 1221 turbulent mixing, and radiative- transfer in the stratocumulus-topped marine boundary- layer
 1222 and comparisons with measurements, *J. Atmos. Sci.*, 52, 1204–1236.
- 1223 Abdul-Razzak, H., and S. J. Ghan (2000), A parameterization of aerosol activation 2. Multiple
 1224 aerosol types, *J. Geophys. Res.*, 105(D5), 6837–6844, doi:10.1029/1999JD901161.
- 1225 Abdul-Razzak, H., S. J. Ghan, and C. Rivera-Carpio (1998), A parameterization of aerosol
 1226 activation - 1. Single aerosol type, *J. Geophys. Res.*, 103(D6), 6123–6131,
 1227 doi:10.1029/97JD03735.
- 1228 Ackerman, T. P., and G. M. Stokes (2003), The Atmospheric Radiation Measurement Program,
 1229 *Phys. Today*, 56, 38–45, doi:10.1063/1.1554135.
- 1230 Ahlgrimm, M., and R. Forbes (2012), The impact of low clouds on surface shortwave radiation
 1231 in the ECMWF model, *Mon. Wea. Rev.*, 140, 3783–3794.
- 1232 Ahlgrimm, M. and R. Forbes (2014), Improving the Representation of Low Clouds and Drizzle
 1233 in the ECMWF Model Based on ARM Observations from the Azores, *Mon. Wea. Rev.*, 142,
 1234 668–685, doi: <http://dx.doi.org/10.1175/MWR-D-13-00153.1>.
- 1235 Anderson, G. P., A. Berk, P. K. Acharya, M. W. Matthew, L. S. Bernstein, J. H. Chetwynd, H.
 1236 Dothe, S. M. Adler-Golden, A. J. Ratkowski, G. W. Felde, J. A. Gardner, M. L. Hoke, S. C.
 1237 Richtsmeier, and L. S. Jeong (2001), MODTRAN4, version 2: Radiative transfer modeling,
 1238 *SPIE-Int. Soc. Opt. Eng.*, 4381, 455– 459.
- 1239 Andreae, M. O. and D. Rosenfeld (2008), Aerosol-cloud-precipitation interactions. Part 1, The
 1240 nature and sources of cloud-active aerosols, *Earth Sci. Rev.*, 89, 13–41.

1241 Bechtold, P., J.-P. Chaboureaud, A. Beljaars, A. K. Betts, M. Köhler, M. Miller, and J.-L.
1242 Redelsperger (2004), The simulation of the diurnal cycle of convective precipitation over
1243 land in a global model, *Quart. J. Roy. Meteor. Soc.*, 130, 3119–3137.

1244 Berg, L. K., and E. I. Kassianov (2008), Temporal Variability of Fair-Weather Cumulus
1245 Statistics at the ACRF SGP Site, *J. Climate*, 21, 3344–3358.

1246 Berg, L. K., E. I. Kassianov, C. N. Long, and D. L. Mills Jr. (2011), Surface summertime
1247 radiative forcing by shallow cumuli at the Atmospheric Radiation Measurement Southern
1248 Great Plains site, *J. Geophys. Res.*, 116, D01202, doi:10.1029/2010JD014593.

1249 Betts, A. K., and C. Jakob (2002), Evaluation of the diurnal cycle of precipitation, surface
1250 thermodynamics, and surface fluxes in the ECMWF model using LBA data, *J. Geophys.*
1251 *Res.*, 107, 8045, doi:10.1029/2001JD000427.

1252 Bony, S., and J.-L. Dufresne (2005), Marine boundary layer clouds at the heart of tropical cloud
1253 feedback uncertainties in climate models, *Geophys. Res. Lett.*, 32, L20806,
1254 doi:10.1029/2005GL023851.

1255 Borque, P., P. Kollias, and S. Giangrande (2014), First Observations of Tracking Clouds Using
1256 Scanning ARM Cloud Radars, *J. Appl. Meteorol. and Climatol.*, early online release
1257 <http://journals.ametsoc.org/doi/abs/10.1175/JAMC-D-13-0182.1>.

1258 Brown, A. R., A. Chlond, C. Golaz, M. Khairoutdinov, D. C. Lewellen, A. P. Lock, M. K.
1259 MacVean, C.-H. Moeng, R. A. J. Neggers, A. P. Siebesma and B. Stevens (2002), Large-eddy
1260 simulation of the diurnal cycle of shallow cumulus convection over land, *Quart. J. Roy. Met.*
1261 *Soc.*, 128, 1075-1094.

1262 Chaboureaud, J.-P., F. Guichard, J.-L. Redelsperger, and J.-P. Lafore (2004), The role of stability
1263 and moisture in the diurnal cycle of convection over land, *Quart. J. Roy. Meteor. Soc.*, 130,
1264 3105–3117.

1265 Chandra, A. S., P. Kollias, and B. A. Albrecht (2013), Multiyear Summertime Observations of
1266 Daytime Fair-Weather Cumuli at the ARM Southern Great Plains Facility, *J. Climate*, 26,
1267 10031–10050, doi: <http://dx.doi.org/10.1175/JCLI-D-12-00223.1>.

1268 Clothiaux, E. E., T. P. Ackerman, G. G. Mace, K. P. Moran, R. T. Marchand, M. A. Miller, and
1269 B. E. Martner (2000), Objective determination of cloud heights and radar reflectivities using
1270 a combination of active remote sensors at the ARM CART sites, *J. Appl. Meteorol.*, 39(5),
1271 645-665.

1272 Davies, L., C. Jakob, K. Cheung, A. DelGenio, A. Hill, T. Hume, R. J. Keane, T. Komori, V. E.
1273 Larson, Y. Lin, B. J. Nielsen, J. Petch, R. S. Plant, M. S. Singh, X. Shi, X. Song, W. Wang,
1274 M. A. Whitall, A. Wolf, S. Xie, G. Zhang (2013), A Single Column Model Ensemble
1275 approach applied to the TWP-ICE experiment, *J. Geophys. Res.*, 118, 6544-6563,
1276 doi:10.1002/jgrd.50450.

1277 Del Genio, A. D., and A. B. Wolf (2000), The temperature dependence of the liquid water path
1278 of low clouds in the southern great planes, *J. Clim.*, 13, 3465–3486, doi:10.1175/1520-
1279 0442(2000)013<3465:TTDOTL>2.0. CO;2.

1280 Dong, X., P. Minnis, and B. Xi (2005), A Climatology of Midlatitude Continental Clouds from
1281 the ARM SGP Central Facility: Part I: Low-Level Cloud Macrophysical, Microphysical, and
1282 Radiative Properties, *J. Climate*, 18, 1391–1410, doi: <http://dx.doi.org/10.1175/JCLI3342.1>.

1283 Dong, X., B. Xi, and P. Minnis (2006), A climatology of midlatitude continental clouds from the
 1284 ARM SGP Central Facility: Part II. Cloud fraction and surface radiative forcing, *J. Clim.*,
 1285 19(5), 1765–1783.

1286 Donner, L. J., B. Wyman, R. S. Hemler, L. W. Horowitz, Y. Ming, M. Zhao, J.-C. Golaz, P.
 1287 Ginoux, S.-J. Lin, M. D. Schwarzkopf, J. Austin, G. Alaka, W. F. Cooke, T. L. Delworth, S.
 1288 Freidenreich, C. T. Gordon, S. M. Griffies, I. M. Held, W. J. Hurlin, S. A. Klein, T. R.
 1289 Knutson, A. R. Langenhorst, H. C. Lee, Y. Lin, B. I. Magi, S. Malyshev, P. C. D. Milly, V.
 1290 Naik, M. J. Nath, R. Pincus, J. J. Ploshay, V. Ramaswamy, C. J. Seman, E. Shevliakova, J. J.
 1291 Sirutis, W. F. Stern, R. J. Stouffer, R. J. Wilson, M. W., A. T. Wittenberg, and F. Zeng
 1292 (2011), The dynamical core, physical parameterizations, and basic simulation characteristics
 1293 of the atmospheric component AM3 of the GFDL Global Coupled Model CM3, *J. Climate*,
 1294 24(13), doi:10.1175/2011JCLI3955.1.

1295 Feng, S., Z. Li, W. Lin, Y. Liu, M. Zhang, T. Toto, A. M. Vogelmann, and S. Endo (2015),
 1296 Development of Fine-Resolution Analyses and Expanded Large-Scale Forcing Properties.
 1297 Part II: Scale-Awareness and Application to Single-Column Model Experiments, *J. Geophys.*
 1298 *Res.*, doi:10.1002/2014JD022254 (in press).

1299 Fridlind, A.M., and A.S. Ackerman (2011), Estimating the sensitivity of radiative impacts of
 1300 shallow, broken marine clouds to boundary-layer aerosol size distribution parameter
 1301 uncertainties for satellite retrieval requirements, *J. Atmos. Ocean. Technol.*, 28, 530-538,
 1302 doi:10.1175/2010JTECHA1520.1.

1303 Fridlind, A.M., A. S. Ackerman, J.-P. Chaboureaud, J. Fan, W. W. Grabowski, A. Hill, T. R.
 1304 Jones, M. M. Khaiyer, G. Liu, P. Minnis, H. Morrison, L. Nguyen, S. Park, J. C. Petch, J.-P.
 1305 Pinty, C. Schumacher, B. Shipway, A. C. Varble, X. Wu, S. Xie, and M. Zhang (2012), A

1306 comparison of TWP-ICE observational data with cloud-resolving model results, *J. Geophys.*
 1307 *Res.*, 117, D05204, doi:10.1029/2011JD016595.

1308 Geoffroy, O., J. Brenguier, and F. Burnet (2010), Parametric representation of the cloud droplet
 1309 spectra for LES warm bulk microphysical schemes, *Atmos. Chem. Phys.*, 10, 4835–4848.

1310 Ghan, S. J., H. Abdul-Razzak, A. Nenes, Y. Ming, X. Liu, M. Ovchinnikov, B. Shipway, N.
 1311 Meskhidze, J. Xu, and X. Shi (2011a), Droplet Nucleation: Physically-based
 1312 Parameterization and Comparative Evaluation, *J. Adv. Model. Earth Syst.*, 3,
 1313 doi:10.1029/2011MS000074.

1314 Ghan, S. J., H. Abdul-Razzak, A. Nenes, Y. Ming, X. Liu, M. Ovchinnikov, B. Shipway, N.
 1315 Meskhidze, J. Xu, and X. Shi (2011b), Correction to “Droplet Nucleation: Physically-based
 1316 Parameterization and Comparative Evaluation,” *J. Adv. Model. Earth Syst.*, 3,
 1317 doi:10.1029/2011MS000107.

1318 Ghate, V. P., B. A. Albrecht, and P. Kollias (2010), Vertical velocity structure of
 1319 nonprecipitating continental boundary layer stratocumulus clouds, *J. Geophys. Res.*, 115,
 1320 D13204, doi:10.1029/2009JD013091.

1321 Goldsmith, J.E.M., F.H. Blair, S.E. Bisson, and D.D. Turner (1998) Turn-key Raman lidar for
 1322 profiling atmospheric water vapor, clouds, and aerosols, *Appl. Opt.*, 37, 4979-4990,
 1323 doi:10.1364/AO.37.004979.

1324 Guichard, F., J. C. Petch, J.-L. Redelsperger, P. Bechtold, J.-P. Chaboureaud, S. Cheinet, W.
 1325 Grabowski, H. Grenier, C. J. Jones, M. Koehler, J.-M. Piriou, R. Tailleux and M. Tomasini
 1326 (2004), Modelling the diurnal cycle of deep precipitating convection over land with CRMs
 1327 and SCMs, *Quart. J. Roy. Meteor. Soc.*, 130, 3139-3172.

1328 Hartmann, D. L., M. E. Ockert-Bell, and M. L. Michelsen (1992), The effect of cloud type on the
1329 earth's energy balance: Global analysis, *J. Climate*, 5, 1281–1304.

1330 Illingworth, A. J., R. J. Hogan, E. J. O'Connor, D. Bouniol, M. E. Brooks, J. Delanoe, D. P.
1331 Donovan, J. D. Eastment, N. Gaussiat, J. W. F. Goddard, M. Haeffelin, H. Klein Baltink, O.
1332 A. Krasnov, J. Pelon, J.-M. Piriou, A. Protat, H. W. J. Russchenberg, A. Seifert, A. M.
1333 Tompkins, G.-J. van Zadelhoff, F. Vinit, U. Willen, D. R. Wilson, and C. L. Wrench (2007),
1334 Cloudnet: Continuous evaluation of cloud profiles in seven operational models using ground-
1335 based observations, *Bull. Amer. Meteor. Soc.*, 88, 883–898.

1336 IPCC (2013), Climate Change 2013: The Physical Science Basis. Contribution of Working
1337 Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change
1338 [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y.
1339 Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United
1340 Kingdom and New York, NY, USA, 1535 pp.

1341 Jiang, H., W. R. Cotton, J. O. Pinto, J. A. Curry, and M. J. Weissbluth (2000), Cloud resolving
1342 simulations of mixed-phase Arctic stratus observed during BASE: Sensitivity to
1343 concentration of ice crystals and large-scale heat and moisture advection, *J. Atmos. Sci.*, 57,
1344 2105–2117.

1345 Kassianov, E., C. N. Long, and M. Ovtchinnikov (2005), Cloud sky cover versus cloud fraction:
1346 Whole-sky simulations and observations, *J. Appl. Meteorol.*, 44, 86–98, doi:10.1175/JAM-
1347 2184.1.

1348 Khairoutdinov, M., and D. Randall (2006), High-resolution simulation of shallow-to-deep
1349 convection transition over land, *J. Atmos. Sci.*, 63, 3421–3436.

1350 Kirkpatrick, M. P., A. S. Ackerman, D. E. Stevens, and N. N. Mansour (2006), On the
 1351 application of the dynamic Smagorinsky model to large-eddy simulations of the cloud-topped
 1352 atmospheric boundary layer, *J. Atmos. Sci.*, 63, 526–546.

1353 Kollias, P., and B. A. Albrecht (2000), The turbulence structure in a continental stratocumulus
 1354 cloud from millimeter-wavelength radar observations, *J. Atmos. Sci.*, 57, 2417–2434,
 1355 doi:10.1175/1520-0469(2000) 057<2417:TTSIAC>2.0.CO;2.

1356 Kollias, P., G. Tselioudis, and B. A. Albrecht (2007), Cloud climatology at the Southern Great
 1357 Plains and the layer structure, drizzle, and atmospheric modes of continental stratus, *J.*
 1358 *Geophys. Res.*, 112, D09116, doi:10.1029/2006JD007307.

1359 Lance, S, A. Nenes, J. Medina, and J.N. Smith (2006), Mapping the operation of the DMT
 1360 Continuous Flow CCN counter, *Aerosol Sci. Technol.*, 40(4), 242-254.

1361 Lenderink, G., A. P. Siebesma, S. Cheinet, S. Irons, C. G. Jones, P. Marquet, F. Müller, D.
 1362 Olmeda, J. Calvo, E. Sánchez, and P. M. M. Soares (2004), The diurnal cycle of shallow
 1363 cumulus clouds over land: A single-column model intercomparison study, *Q.J.R. Meteorol.*
 1364 *Soc.*, 130: 3339–3364. doi: 10.1256/qj.03.122.

1365 Li, Z., Y. Chao, and J.D. Farrara, and J.C. McWilliams (2012), Impacts of distinct observations
 1366 during the 2009 Prince William Sound field experiment: A data assimilation study, *Cont.*
 1367 *Shelf Res.*, DOI:10.1016/j.csr.2012.06.018.

1368 Li, Z., S. Feng, Y. Liu, W. Lin, M. Zhang, T. Toto, A. M. Vogelmann, and S. Endo (2014),
 1369 Development of Fine-Resolution Analyses and Expanded Large-Scale Forcing Properties.
 1370 Part I: Methodology and Evaluation, *J. Geophys. Res.*, (accepted), doi:
 1371 10.1002/2014JD022245.

1372 Liu, G., Y. Liu, and S. Endo (2013), Evaluation of Surface Flux Parameterizations with Long-
 1373 Term ARM Observations, *Mon. Weather Rev.*, 141, 773-797, doi:10.1175/MWR-D-12-
 1374 00095.1.

1375 Lu, C., Y. Liu, S. Niu, and A. M. Vogelmann (2012a), Lateral entrainment rate in shallow
 1376 cumuli: Dependence on dry air sources and probability density functions, *Geophys. Res.*
 1377 *Lett.*, 39, L20812, doi:10.1029/2012GL053646.

1378 Lu, C., Y. Liu, S. Niu, and A. M. Vogelmann (2012b), Observed impacts of vertical velocity on
 1379 cloud microphysics and implications for aerosol indirect effects, *Geophys. Res. Lett.*, 39,
 1380 L21808, doi:10.1029/2012GL053599.

1381 Lu, C., S. Niu, Y. Liu, and A. M. Vogelmann (2013), Empirical relationship between
 1382 entrainment rate and microphysics in cumulus clouds, *Geophys. Res. Lett.*, 40, 2333–2338,
 1383 doi:10.1002/grl.50445.

1384 Lu, C., Y. Liu, S. Niu, and S. Endo (2014), Scale dependence of entrainment-mixing
 1385 mechanisms in cumulus clouds, *J. Geophys. Res. Atmos.*, 119, 13,877–13,890,
 1386 doi:10.1002/2014JD022265.

1387 Markwardt, C. B. (2009), Non-Linear Least Squares Fitting in IDL with MPFIT, in proc.
 1388 *Astronomical Data Analysis Software and Systems XVIII*, Quebec, Canada, ASP Conference
 1389 Series, Vol. 411, eds. D. Bohlender, P. Dowler & D. Durand (Astronomical Society of the
 1390 Pacific: San Francisco), p. 251-254 (ISBN: 978-1-58381-702-5).

1391 Mather, J. H., and J. W. Voyles (2013), The Arm Climate Research Facility: A Review of
 1392 Structure and Capabilities, *Bull. Amer. Meteor. Soc.*, 94, 377–392,
 1393 doi: <http://dx.doi.org/10.1175/BAMS-D-11-00218.1>

1394 McFarquhar, G.M., S. Ghan, J. Verlinde, A. Korolev, J. W. Strapp, B. Schmid, J. M. Tomlinson,
 1395 M. Wolde, S. D. Brooks, D. Cziczo, M. K. Dubey, J. Fan, C. Flynn, I. Gultepe, J. Hubbe, M.
 1396 K. Gilles, A. Laskin, P. Lawson, W. R. Leaitch, P. Liu, X. Liu, D. Lubin, C. Mazzoleni, A.-
 1397 M. Macdonald, R. C. Moffet, H. Morrison, M. Ovchinnikov, M. D. Shupe, D. D. Turner, S.
 1398 Xie, A. Zelenyuk, K. Bae, M. Freer, and A. Glen (2011), Indirect and Semi-Direct Aerosol
 1399 Campaign (ISDAC): The impact of arctic aerosols on clouds, *Bull. Amer. Meteor. Soc.*, 92,
 1400 183-201, doi:10.1175/2010BAMS2935.1.

1401 Mechem, D. B., Y. L. Kogan, and D. M. Schultz (2010), Large-eddy observation of post-cold-
 1402 frontal continental stratus, *J. Atmos. Sci.*, 67, 3835–3853.

1403 Mei, F., P. L. Hayes, A. M. Ortega, J. W. Taylor, J. D. Allan, J. B. Gilman, W. C. Kuster, J. A.
 1404 de Gouw, J. L. Jimenez, and J. Wang (2013a), Droplet activation properties of organic
 1405 aerosols observed at an urban site during CalNex-LA, *J. Geophys. Res.*, 118, 2903-2917,
 1406 doi: 10.1002/jgrd.50285.

1407 Mei, F., A. Setyan, Q. Zhang, and J. Wang (2013b), CCN activity of organic aerosols observed
 1408 downwind of urban emissions during CARES, *Atmos. Chem. Phys.*, 13, 12155-12169,
 1409 doi: 10.5194/acp-13-12155-2013.

1410 Melfi, S. H., K. D. Evans, J. Li, D. Whiteman, R. Ferrare, and G. Schwemmer (1997),
 1411 Observation of Raman scattering by cloud droplets in the atmosphere, *Appl. Opt.*, 36, 3551-
 1412 3559, <http://dx.doi.org/10.1364/AO.36.003551>.

1413 Moore, R.H., R. Bahreini, C.A. Brock, K.D. Froyd, J. Cozic, J.S. Holloway, A.M. Middlebrook,
 1414 D.M. Murphy, and A. Nenes (2011), Hygroscopicity and Composition of Alaskan Arctic
 1415 CCN During April 2008, *Atmos. Chem. Phys.*, 11, 11807-11825, doi:10.5194/acp-11-11807-
 1416 2011.

1417 Morrison, H., and W. W. Grabowski (2008), Modeling supersaturation and subgrid-scale mixing
 1418 with two-moment bulk warm microphysics, *J. Atmos. Sci.*, 65, 792–812,
 1419 doi:10.1175/2007JAS2374.1.

1420 Morrison, H., J. Curry, and V. Khvorostyanov (2005), A new double-moment microphysics
 1421 parameterization for application in cloud and climate models. Part I: Description, *J. Atmos.*
 1422 *Sci.*, 62, 1665–1677, doi: <http://dx.doi.org/10.1175/JAS3446.1>.

1423 Morrison, H., P. Zuidema, A.S. Ackerman, A. Avramov, G. de Boer, J. Fan, A.M. Fridlind, T.
 1424 Hashino, J.Y. Harrington, Y. Luo, M. Ovchinnikov, and B. Shipway (2011), Intercomparison
 1425 of cloud model simulations of Arctic mixed-phase boundary layer clouds observed during
 1426 SHEBA/FIRE-ACE, *J. Adv. Model. Earth Syst.*, 3, M06003, doi:10.1029/2011MS000066.

1427 Neale, R. B., C.-C. Chen, A. Gettelman, P. H. Lauritzen, S. Park, D. L. Williamson, A. J.
 1428 Conley, R. Garcia, D. Kinnison, J.-F. Lamarque, D. Marsh, M. Mills, A. K. Smith, S. Tilmes,
 1429 F. Vitt, H. Morrison, P. Cameron-Smith, W. D. Collins, M. J. Iacono, R. C. Easter, S. J.
 1430 Ghan, X. Liu, P. J. Rasch, and M. A. Taylor (2012), Description of the NCAR Community
 1431 Atmosphere Model (CAM 5.0), NCAR Technical Note, NCAR/TN-486+STR.

1432 Neggers, R.A.J., A.P. Siebesma, and T. Heus (2012), Continuous single-column model
 1433 evaluation at a permanent meteorological supersite, *Bull. Amer. Meteorol. Soc.*, 93, 1389-
 1434 1400, 93, 1389–1400, doi:10.1175/BAMS-D-11-00162.1.

1435 Newsom, R.K., D.D. Turner, and J.E.M. Goldsmith (2013), Long-term evaluation of temperature
 1436 profiles measured by an operational Raman lidar, *J. Atmos. Oceanic Technol.*, 30, 1616-
 1437 1634, doi:10.1175/JTECH-D-12-00138.1.

1438 Parworth, C., J. Fast, F. Mei, T. Shippert, C. Sivaraman, A. Tilp, T. Watson, and Q. Zhang
 1439 (2015), Long-term Measurements of Submicrometer Aerosol Chemistry at the Southern

1440 Great Plains (SGP) Using an Aerosol Chemical Speciation Monitor (ACSM), *Atmos.*
 1441 *Environ.*, 106, 43-55, doi:10.1016/j.atmosenv.2015.01.060
 1442 Petters, M. D., and S. M. Kreidenweis (2007), A single parameter representation of hygroscopic
 1443 growth and cloud condensation nucleus activity, *Atmos. Chem. Phys.*, 7, 1961–1971,
 1444 doi: 10.5194/acp-7-1961-2007.
 1445 Pringle, K. J., H. Tost, A. Pozzer, U. Pöschl, and J. Lelieveld (2010), Global distribution of the
 1446 effective aerosol hygroscopicity parameter for CCN activation, *Atmos. Chem. Phys.*, 10,
 1447 5241–5255, doi:10.5194/acp-10-5241-2010.
 1448 Randall, D., and D. Cripe (1999), Alternative methods for specification of observed forcing in
 1449 single-column models and cloud system models, *J. Geophys. Res.*, 104(D20), 24527-24545,
 1450 doi: 10.1029/1999JD900765.
 1451 Rio, C., F. Hourdin, J.-Y. Grandpeix, and J.-P. Lafore (2009), Shifting the diurnal cycle of
 1452 parameterized deep convection over land, *Geophys. Res. Lett.*, 36, L07809,
 1453 doi:10.1029/2008GL036779.
 1454 Rochetin, N., F. Couvreux, J.-Y. Grandpeix, and C. Rio (2014), Deep Convection Triggering by
 1455 Boundary Layer Thermals. Part I: LES Analysis and Stochastic Triggering Formulation, *J.*
 1456 *Atmos. Sci.*, 71, 496–514, doi: <http://dx.doi.org/10.1175/JAS-D-12-0336.1>.
 1457 Rose, D., A. Nowak, P. Achtert, A. Wiedensohler, M. Hu, M. Shao, Y. Zhang, M. O. Andreae,
 1458 and U. Pöschl (2010), Cloud condensation nuclei in polluted air and biomass burning smoke
 1459 near the mega-city Guangzhou, China – Part 1: Size-resolved measurements and implications
 1460 for the modeling of aerosol particle hygroscopicity and CCN activity, *Atmos. Chem. Phys.*,
 1461 10, 3365–3383, doi:10.5194/acp-10-3365-2010.

1462 Schmid, B., J. M. Tomlinson, J. M. Hubbe, J. M. Comstock, F. Mei, D. Chand, M. S. Pekour, C.
 1463 D. Kluzek, E. Andrews, S. C. Biraud, and G. M. McFarquhar (2014), The DOE ARM Aerial
 1464 Facility, *Bull. Amer. Meteor. Soc.*, 95 (5), 723-742, doi: [http://dx.doi.org/10.1175/BAMS-D-](http://dx.doi.org/10.1175/BAMS-D-13-00040.1)
 1465 13-00040.1.

1466 Schmidt, G.A., R. Ruedy, J.E. Hansen, I. Aleinov, N. Bell, M. Bauer, S. Bauer, B. Cairns, V.
 1467 Canuto, Y. Cheng, A. Del Genio, G. Faluvegi, A.D. Friend, T.M. Hall, Y. Hu, M. Kelley,
 1468 N.Y. Kiang, D. Koch, A.A. Lacis, J. Lerner, K.K. Lo, R.L. Miller, L. Nazarenko, V. Oinas,
 1469 Ja. Perlwitz, Ju. Perlwitz, D. Rind, A. Romanou, G.L. Russell, Mki. Sato, D.T. Shindell, P.H.
 1470 Stone, S. Sun, N. Tausnev, D. Thresher, and M.-S. Yao (2006), Present day atmospheric
 1471 simulations using GISS ModelE: Comparison to in-situ, satellite and reanalysis data, *J.*
 1472 *Climate*, 19, 153-192, doi:<http://dx.doi.org/10.1175/JCLI3612.1>.

1473 Sengupta, M., E. E. Clothiaux, T. P. Ackerman, S. Kato, and Q. Min (2003), Importance of
 1474 accurate liquid water path for estimation of solar radiation in warm boundary layer clouds:
 1475 An observational study, *J. Climate*, 16, 2997–3009, doi:[http://dx.doi.org/10.1175/1520-](http://dx.doi.org/10.1175/1520-0442(2003)016<2997:IOALWP>2.0.CO;2)
 1476 [0442\(2003\)016<2997:IOALWP>2.0.CO;2](http://dx.doi.org/10.1175/1520-0442(2003)016<2997:IOALWP>2.0.CO;2)

1477 Slingo, A., K. I. Hodges, and G. J. Robinson (2004), Simulation of the diurnal cycle in a climate
 1478 model and its evaluation using data from METEOSAT 7, *Q. J. R. Meteorol. Soc.*, 130, 1449–
 1479 1467, doi:10.1256/ qj.03.165.

1480 Sherwood, S. C., S. Bony, and J.-L. Dufresne (2014), Spread in model climate sensitivity traced
 1481 to atmospheric convective mixing, *Nature*, 505, 37–42, doi:10.1038/nature12829.

1482 Stevens, D. E., A. S. Ackerman, and C. S. Bretherton (2002), Effects of domain size and
 1483 numerical resolution on the simulation of shallow cumulus convection, *J. Atmos. Sci.*, 59,
 1484 3285–3301, doi:[http://dx.doi.org/10.1175/1520-0469\(2002\)059<3285:EODSAN>2.0.CO;2](http://dx.doi.org/10.1175/1520-0469(2002)059<3285:EODSAN>2.0.CO;2).

1485 Stokes, G. M., and S. E. Schwartz (1994), The Atmospheric Radiation Measurement (ARM)
 1486 program: Programmatic background and design of the cloud and radiation testbed, *Bull.*
 1487 *Amer. Meteor. Soc.*, 75, 1201-1221, doi:[http://dx.doi.org/10.1175/1520-](http://dx.doi.org/10.1175/1520-0477(1994)075<1201:TARMPP>2.0.CO;2)
 1488 [0477\(1994\)075<1201:TARMPP>2.0.CO;2](http://dx.doi.org/10.1175/1520-0477(1994)075<1201:TARMPP>2.0.CO;2)
 1489 Toon, O. B., C. P. McKay, T. P. Ackerman, and K. Santhaman (1989), Rapid calculation of
 1490 radiative heating and photodissociation rates in inhomogeneous multiple scattering
 1491 atmospheres, *J. Geophys. Res.*, 94, 16287–16301.
 1492 Turner, D.D. (2007), Improved ground-based liquid water path retrievals using a combined
 1493 infrared and microwave approach, *J. Geophys. Res.*, 112, D15204,
 1494 doi:10.1029/2007JD008530.
 1495 Turner, D.D., R.A. Ferrare, L.A. Heilman Brasseur, W.F. Feltz, and T.P. Tooman (2002),
 1496 Automated retrievals of water vapor and aerosol profiles from an operational Raman lidar, *J.*
 1497 *Atmos. Oceanic Technol.*, 19, 37-50, doi:10.1175/1520-0426(2002)019<0037:AROWVA>2.0.CO;2.
 1498
 1499 Turner, D.D., S.A. Clough, J.C. Liljegren, E.E. Clothiaux, K. Cady-Pereira, and K.L. Gaustad
 1500 (2007a), Retrieving liquid water path and precipitable water vapor from Atmospheric
 1501 Radiation Measurement (ARM) microwave radiometers, *IEEE Trans. Geosci. Remote Sens.*,
 1502 45, 3680-3690, doi:10.1109/TGRS.2007.903703.
 1503 Turner, D.D, A. M. Vogelmann, R. Austin, J.C. Barnard, K. Cady-Pereira, C. Chiu, S.A. Clough,
 1504 C.J. Flynn, M.M. Khaiyer, J.C. Liljegren, K. Johnson, B. Lin, C.N. Long, A. Marshak, S.Y.
 1505 Matrosov, S.A. McFarlane, M.A. Miller, Q. Min, P. Minnis, W. O'Hirok, Z. Wang, and W.
 1506 Wiscombe (2007b), Thin liquid water clouds: Their importance and our challenge, *Bull.*
 1507 *Amer. Meteor. Soc.*, 88,177-190, doi:<http://dx.doi.org/10.1175/BAMS-88-2-177>.

1508 Turner, D.D., R.A. Ferrare, V. Wulfmeyer, and A.J. Scarino (2014a), Aircraft evaluation of
 1509 ground-based Raman lidar water vapor turbulence profiles in convective mixed layers, *J.*
 1510 *Atmos. Oceanic Technol.*, 31, 1078-1088, doi:10.1175/JTECH-D-13-00075-1.
 1511 Turner, D. D., V. Wulfmeyer, L. K. Berg, and J. H. Schween (2014b), Water vapor turbulence
 1512 profiles in stationary continental convective mixed layers, *J. Geophys. Res. Atmos.*, 119,
 1513 doi:10.1002/2014JD022202.
 1514 Vogelmann, A. M. (2012), *RACORO Data guide v1*, DOE/SC-ARM-10-031, 23 pp
 1515 (<http://www.arm.gov/publications/programdocs/doe-sc-arm-10-031.pdf?id=85>).
 1516 Vogelmann, A. M., G. M. McFarquhar, J. A. Ogren, D. D. Turner, J. M. Comstock, G. Feingold,
 1517 C. N. Long, H. H. Jonsson, A. Bucholtz, D. R. Collins, G. S. Diskin, H. Gerber, R. P.
 1518 Lawson, R. K. Woods, E. Andrews, H.-J. Yang, J. C. Chiu, D. Hartsock, J. M. Hubbe, C. Lo,
 1519 A. Marshak, J. W. Monroe, S. A. McFarlane, B. Schmid, J. M. Tomlinson, and T. Toto
 1520 (2012), RACORO Extended-Term, Aircraft Observations of Boundary-Layer Clouds, *Bull.*
 1521 *Amer. Meteor. Soc.*, 93, 861–878, doi:<http://dx.doi.org/10.1175/BAMS-D-11-00189.1>
 1522 Walker, J. T., D. R. Whitall, W. Robarge, and H. W. Paerl (2004), Ambient ammonia and
 1523 ammonium aerosol across a region of variable ammonia emission density, *Atmos. Environ.*,
 1524 38, 1235–1246, doi:10.1016/j.atmosenv.2003.11.027.
 1525 Wang, J., Y. N. Lee, P. H. Daum, J. Jayne, and M. L. Alexander (2008), Effects of aerosol
 1526 organics on cloud condensation nucleus (CCN) concentration and first indirect aerosol effect,
 1527 *Atmos. Chem. Phys.*, 8, 6325-6339, doi:10.5194/acpd-8-9783-2008.
 1528 Wu, W., Y. Liu, M. P. Jensen, T. Toto, M. J. Foster, and C. N. Long (2014), A comparison of
 1529 multiscale variations of decade-long cloud fractions from six different platforms over the

1530 Southern Great Plains in the United States, *J. Geophys. Res. Atmos.*, 119, 3438–3459,
 1531 doi:10.1002/2013JD019813.

1532 Wulfmeyer, V., S. Pal, D.D. Turner, and E. Wagner (2010), Can water vapour Raman lidar
 1533 resolve profiles of turbulent variables in the convective boundary layer? *Boundary Layer*
 1534 *Meteor.*, 136, 253-284, doi:10.1007/s10546-010-9494-z.

1535 Xie, S., R. T. Cederwall, and M. Zhang (2004), Developing long-term single-column
 1536 model/cloud system-resolving model forcing data using numerical weather prediction
 1537 products constrained by surface and top of the atmosphere observations, *J. Geophys. Res.*,
 1538 109(D1), D01104, doi:10.1029/2003JD004045

1539 Xie, S., R. B. McCoy, S. A. Klein, R. T. Cederwall, W. J. Wiscombe, M. P. Jensen, K. L.
 1540 Johnson, E. E. Clothiaux, K. L. Gaustad, C. N. Long, J. H. Mather, S. A. McFarlane, Y. Shi,
 1541 J.-C. Golaz, Y. Lin, S. D. Hall, R. A. McCord, G. Palanisamy, and D. D. Turner (2010),
 1542 CLOUDS AND MORE: ARM Climate Modeling Best Estimate Data, *Bull. Amer. Meteor.*
 1543 *Soc.*, 91, 13–20, doi:10.1175/2009BAMS2891.1.

1544 Zhang, M. H., and J. L. Lin (1997), Constrained variational analysis of sounding data based on
 1545 column-integrated budgets of mass, heat, moisture, and momentum: Approach and
 1546 application to ARM measurements, *J. Atmos. Sci.*, 54(11), 1503-1524,
 1547 doi:[http://dx.doi.org/10.1175/1520-0469\(1997\)054<1503:CVAOSD>2.0.CO;2](http://dx.doi.org/10.1175/1520-0469(1997)054<1503:CVAOSD>2.0.CO;2)

1548 Zhang, M. H., J. L. Lin, R. T. Cederwall, J. J. Yio, and S. C. Xie (2001), Objective Analysis of
 1549 ARM IOP Data: Method and Sensitivity, *Mon. Weather Rev.*, 129(2), 295-311,
 1550 doi:[http://dx.doi.org/10.1175/1520-0493\(2001\)129<0295:OAOAID>2.0.CO;2](http://dx.doi.org/10.1175/1520-0493(2001)129<0295:OAOAID>2.0.CO;2)

1551 Zhang, M., C.S. Bretherton, P.N. Blossey, P.H. Austin, J.T. Bacmeister, S. Bony, F. Brient, S.K.
 1552 Cheedela, A. Cheng, A.D. Del Genio, S.R. De Roode, S. Endo, C.N. Franklin, J.-C. Golaz,

1553 C. Hannay, T. Heus, F.A. Isotta, J.-L. Dufresne, I.-S. Kang, H. awai, M. Köehler, V.E.
 1554 Larson, Y. Liu, A.P. Lock, U. Lohmann, M.F. Khairoutdinov, A.M. Molod, R.A.J. Neggers,
 1555 P. Rasch, I. Sandu, R. Senkbeil, A.P. Siebesma, C. Siegenthaler-Le Drian, B. Stevens, M.J.
 1556 Suarez, K.-M. Xu, K. von Salzen, M.J. Webb, A. Wolf, and M. Zhao (2013), CGILS: Results
 1557 from the first phase of an international project to understand the physical mechanisms of low
 1558 cloud feedbacks in single column models, *J. Adv. Model. Earth Syst.*, 5, 826-842,
 1559 doi:10.1002/2013MS000246.

1560 Zhang, Y., and S. A. Klein (2010), Mechanisms affecting the transition from shallow to deep
 1561 convection over land: Inferences from observations of the diurnal cycle collected at the ARM
 1562 Southern Great Plains site, *J. Atmos. Sci.*, 67, 2943– 2959,
 1563 doi:<http://dx.doi.org/10.1175/2010JAS3366.1>

1564 Zhang, Y., and S. A. Klein (2013), Factors controlling the vertical extent of fair-weather shallow
 1565 cumulus clouds over land: Investigation from diurnal-cycle observations of the diurnal cycle
 1566 collected at the ARM Southern Great Plains site, *J. Atmos. Sci.*, 70, 1297–1315,
 1567 doi:10.1175/JAS-D-12-0131.1.

1568

TABLE AND FIGURE CAPTIONS

1569 **Table 1. Kappa values per flight.** A single κ value is obtained from measurements by
 1570 simultaneously fitting the CCN(SS) and the aerosol size distributions for the beginning (spiral 1)
 1571 and end (spiral 2) of each flight. See text for details. No spiral was flown at the start of the
 1572 flight on 6 May, indicated by ‘---’.

1573 **Table A1. RACORO airborne instrumentation.** The symbol \uparrow means upward-looking and \downarrow
 1574 means downward-looking. The measurement rates given (e.g., 10 Hz) represent the upper limits
 1575 possible; the data might be available at lower rates.

1576 **Figure 1. RACORO campaign cloud statistics.** Boxes are medians and bars extend between
 1577 the 25th and 75th percentiles. Colors indicate the dominant cloud type as given in the legend
 1578 (from *Vogelmann et al.* [2012]). The selected days are denoted by the dates (in May) given
 1579 below their medians (black-filled squares), and the case number is denoted by the symbols
 1580 encircling the date, where box=case 1 (all Cu), triangle=case 2 (Cu and St), circle=case 3 (Sc,
 1581 Sc-Cu and St). (a) Microscale properties of cloud-droplet number concentration (N_d) and CCN
 1582 at 0.2% supersaturation, $CCN_{0.2\%}$. $CCN_{0.2\%}$ are from all on-station measurements outside of
 1583 cloud ($LWC < 0.01 \text{ g m}^{-3}$), which are concentrated at altitudes near the cloud layer. (b)
 1584 Macroscale properties of cloud fraction and cloud LWP. Cloud fraction is from the ARSCL
 1585 cloud mask of overhead cloud frequency using a 20-min moving average. LWP is the product of
 1586 cloud thickness, from ARSCL, and LWC, from in-situ CAS measurements (see text). The LWP
 1587 median and percentiles are products of the respective LWC and cloud thickness percentiles.
 1588 Eight flights are not plotted in (b) because they were not coincident with the ARSCL
 1589 measurements.

Figure 2. Three RACORO case study periods. Shown are the 60-h case study periods: a) case 1 cumulus, 22-24 May; b) case 2 cumulus and drizzling stratus followed by clear sky, 26-28 May; and c) variable cloud types, 6-8 May (see text for details). The bottom axis is time as decimal day in May (UTC) and the top axis is local solar time (UTC minus 6 h); vertical dashed lines indicate solar noon. The time-height ARSCL cloud fractional occurrence per 10-min interval is given for each case, and flight periods are indicated by green boxes. Below each period is a representative hemispheric view of the cloud field from the Total-Sky Imager (TSI) at approximately the middle of the flight. The radial band in the TSI images is the sun-blocking band. No flight was flown on the last day of case 2, 28 May. Note the enhanced resolution at the bottom of the ARSCL color scale.

Figure 3. Representative example of three-mode aerosol size distribution fits. Data are from the second spiral on 23 May for the 100-m height-interval centered on 1,150 m above mean sea-level (AMSL), or about 800 m above ground level. The black line is the combined average of the SMPS and PCASP observations. The red dashed curve is the best fit obtained by allowing the geometric mean diameter, D , and geometric standard deviation, σ , to vary independently with altitude bin (D values indicated by vertical red lines). The blue curve is the resulting fit when D and σ are fixed per spiral profile and the number concentrations per mode are refit per height interval (fixed D values indicated by the vertical blue lines). The quality of the fit degrades slightly when D and σ are fixed but it still provides good overall agreement, particularly for the particle diameters greater than $\sim 0.1 \mu\text{m}$ that are most active as CCN for supersaturations $\leq 0.6\%$ in our case studies.

Figure 4. Example of κ fitting. The observed CCN supersaturation spectrum, CCN(SS), is the average of the observations (i.e., averages of the 60 s 1-Hz averages) from 22 May (solid black

line), and standard errors are given for each supersaturation (0.19, 0.23, 0.30, 0.40, 0.53). The same averaged CCN data are used in separate κ analyses with the aerosol fits from the spiral at the start (spiral 1) and end of the flight (spiral 2). The κ values determined are consistent between the spirals being, respectively, 0.08 and 0.07.

Figure 5. Sensitivity to advection and surface forcing components. Observations and DHARMA trial simulations are shown for the three case study periods in terms of lifting condensation level height (Z_{LCL}) and mid-boundary layer average (300-500 m) water vapor mixing ratio ($\overline{q_v}$), temperature (\overline{T}), and relative humidity (\overline{RH}). Local solar time (LST) is indicated at the top of the figure and decimal day (UTC) in May at the bottom. A one-hour moving average is applied to the observations, as described in Appendix B. Observations and their uncertainties are represented by, respectively, black lines and gray shading. Simulations are for the ARM standard domain without relaxation (ARM_D, orange), for surface forcing as in ARM_D but with no large-scale forcing (ARM_D-noLS, blue), and large-scale forcing as in ARM_D but with the domain-averaged EBBR surface fluxes replaced by ECOR values at the SGP central facility (ARM_D-SF, red). Yellow vertical bars indicate the aircraft flight periods. Simulations are not plotted for case 3 after 8.5 May, as discussed in the text.

Figure 6. Cloud fraction profile sensitivity. Shown are cloud fraction profiles observed by ARSCL and from simulations for the three case studies periods for ARM_D, ARM_D-SF, and ARM_D-noLS. Local solar time (LST) is indicated at the top and decimal day (UTC) in May at the bottom. A two-toned vertical scale is used, where the vertical region below 4 km is expanded and above 4 km is reduced; the partition between the regions is indicated by a dashed line. This enables viewing the full tropospheric cloud profiles without sacrificing details of the boundary layer clouds, as would be the case if a single linear scale were used. Yellow vertical

bars indicate the aircraft flight periods. Simulations are not plotted for case 3 after 8.5 May, as discussed in the text.

Figure 7. Simulated atmospheric state for ensemble forcings. As for Figure 5, but for DHARMA simulations using the ARM forcing for the standard 300-km domain (ARM_D, orange) and 150-km reduced domain (ARM_d, red), and for the ECMWF reduced domain (ECMWF_d, blue). Reduced-domain MS-DA simulations (MS-DA_d, green) are run using WRF (see text). Dashed lines use 12-h thermodynamic relaxation, and solid lines are without relaxation. Yellow vertical bars indicate the aircraft flight periods. (Note that case 2 simulations end at 28.8 May due to WRF stability issues.)

Figure 8. Cloud-mask profile for ensemble forcings. 2D cloud masks show the time-height location of cloud (not cloud fraction) for ARSCL observations and DHARMA simulations with and without 12-h thermodynamic relaxation for ARM_D, ARM_d, and ECMWF_d. The MS-DA_d simulations are run using WRF. Green indicates where the simulations with and without relaxation both have cloud, red is where cloud is only present without relaxation, and blue is where cloud is only present with relaxation. The same two-toned vertical scale is used as in Fig. 6, where the vertical region below 4 km is expanded and above 4 km is reduced. Yellow vertical bars indicate the aircraft flight periods.

Figure 9. Cloud fraction and \overline{LWP} for ensemble forcings. Observed and simulated domain-averaged cloud fraction and \overline{LWP} for the three cases (columns) using DHARMA for ARM_D, ARM_d, ECMWF_d, and WRF for MS-DA_d. Shown are cloud fraction without thermodynamic relaxation (row 1), cloud fraction with relaxation (row 2), \overline{LWP} without relaxation (row 3), and \overline{LWP} with relaxation (row 4). Cloud fraction observations are from the

TSI (black line) and ARSCL (dark gray line) with shading in between when both are present. \overline{LWP} observations are from MIXCRA during the daytime and MWRRet during the nighttime; when MWRRet is used its uncertainty is indicated by gray shading below 30 g m^{-2} . Observed cloud fraction and \overline{LWP} use one-hour smoothing for clarity. Yellow vertical bars indicate the aircraft flight periods. Note that simulated cloud fraction and \overline{LWP} are computed for cloud < 8 km.

Figure 10. Initial condition RH profiles. The lower-atmospheric RH profiles (< 3.5 km) are shown from the observed soundings used as initial conditions for the three cases: case 1 sounding for 22 May 11:30 UTC (red), case 2 sounding for 26 May 11:28 UTC (blue), and case 3 sounding for 6 May 11:27 UTC (green). Their respective Z_{LCL} values are given as horizontal dashed lines.

Figure 11. Flight period relationship between mean cloud fraction and \overline{LWP} . Flights for the three cases are given per row. A point is an average over each flight period of the domain cloud fraction and domain-mean \overline{LWP} computed from the simulations (see legend); pluses are with thermodynamic relaxation and squares are without it. The horizontal extent of the black observation bars represents the cloud fraction range from the TSI and ARSCL estimates using a 20-min moving average (as in Fig. 1b). The vertical bar—when visible—is the MIXCRA retrieval uncertainty after averaging across the domain, where the cloud-free measurement uncertainty is effectively zero. Observations are circled for clarity in panels b, c, and e. The gray bar in 5/22 estimates the values excluding the early morning residual layer (see text). Results for the 5/8 flight are not plotted because most simulations are invalid after an earlier

1679 passage of a squall line to the north of the central facility (see text). MS-DA without relaxation
1680 is not plotted in (b) because it is significantly off scale (cloud fraction = 0.87, \overline{LWP} =199 g m⁻²).

1681

TABLES AND FIGURES

1682 **Table 1. Kappa values per flight.** A single κ value is obtained from measurements by
 1683 simultaneously fitting the CCN(SS) and the aerosol size distributions for the beginning (spiral 1)
 1684 and end (spiral 2) of each flight. See text for details. No spiral was flown at the start of the
 1685 flight on 6 May, indicated by '---'.

1686

1687

Case	May Date	Kappa (spiral 1, spiral 2)
1	22	0.08, 0.07
	23	0.10, 0.11
	24	0.11, 0.12
2	26	0.11, 0.11
	27	0.13, 0.12
3	6	----, 0.10
	7	0.04, 0.04
	8	0.11, 0.11

1688 **Table A1. RACORO airborne instrumentation.** The symbol ↑ means upward-looking and ↓ means downward-looking. The
1689 measurement rates given (e.g., 10 Hz) represent the upper limits possible; the data might be available at lower rates.

		MEASUREMENT	INSTRUMENT	SPECIFICATIONS AND/OR COMMENTS
CATEGORY	Cloud Microphysics	Liquid-Water Content (LWC)	Particle Volume Monitor-100A (Gerber Probe)	LWC and Effective radius; 100 Hz reported at 10 Hz LWC sensitivity roll off starts ~30 μm
			SEA LWC Probe (WCM-2000, LWC only)	10 Hz
		Drop Size Distribution (in diameter)	Forward Scattering Spectrometer Probe-100 (FSSP)	2–30 μm at 1 Hz
			Cloud, Aerosol Precipitation Spectrometer (CAPS)	0.5–1,550 μm; Consists of a CAS (0.5–50 μm) at 10 Hz and a 1D CIP (25–1,550 μm) at 1 Hz
			2D Cloud Imaging Probe (2D CIP)	50–1,600 μm, at 1 Hz
			2D Stereo Probe (2D-S)	10–1,280 μm, 10 Hz reported at 1 Hz
		Cloud Extinction	Cloud Integrating Nephelometer (CIN)	Extinction at 100 Hz, averaged to 10 Hz
	Radiation	Broadband Irradiances	↑↓ Shortwave Kipp & Zonen (a modified CM22; BBSR)	0.2 Hz for 95% response, logged at 100 Hz and stored at 10 Hz; no dome/sink temperatures
			↑↓ Longwave Kipp & Zonen (a modified CG4; BBIR)	Same as for shortwave
			↑ Sunshine Pyranometer (SPN1)	Direct-diffuse partitioning; 3–5 Hz for 95% response, logged at 100 Hz and stored at 10 Hz
		Spectral Irradiances	↑↓ Multifilter Radiometer (MFR)	5-channels 415-867 nm with 1625 nm at 10 Hz
			↑↓ HydroRad-3 Hyperspectral Radiometer	350–850nm, 0.3-2.5 nm resolution every 1–6 sec
		Spectral Radiances	↑ or ↓ HydroRad-3 Hyperspectral Radiometer	3° FOV, 350-850 nm, 0.3–2.5 nm res. every 1–6 sec
			↑↓ Infrared Thermometer (IRT)	10 Hz
	Aerosol	Cloud Condensation Nuclei (CCN)	Dual-Column CCN Spectrometer	Constant 0.2% supersaturation (SS) at 1 Hz Full SS scan in ~25 min (0.8%, 0.57, 0.4, 0.28, 0.2)
		Size Distribution (in diameter)	Two Condensation Nuclei Particle Counters (CPCs)	D > 10 nm (CPC1), and D > 15 nm (CPC2) at 1 Hz
			Ultrafine CPC (UPC or UFCPC)	D > 3 nm at 1 Hz
			Scanning Mobility Particle Sizer (SMPS)	D from 12–600 nm every ~60 secs
			Passive Cavity Aerosol Spectrometer Probe (PCASP)	approximately 100–2,200 nm at 1 Hz
	Atmospheric State	Temperature	Rosemount and Vaisala (backup)	100 Hz; Measurement uncertainty inside of cloud
		Water Vapor	2 Chilled Mirror Hygrometers (EdgeTech, CR2)	~1 Hz for T > -40°C
			Diode Laser Hygrometer (DLH)	100 Hz, No equilibration needed after leaving cloud
		Horizontal wind & Vertical velocity	Determined from multiple aircraft sensors	10 Hz
		Conditions	Handheld photos and DAQ flight images/video	DAQ images every 1–2 sec; forward and side view

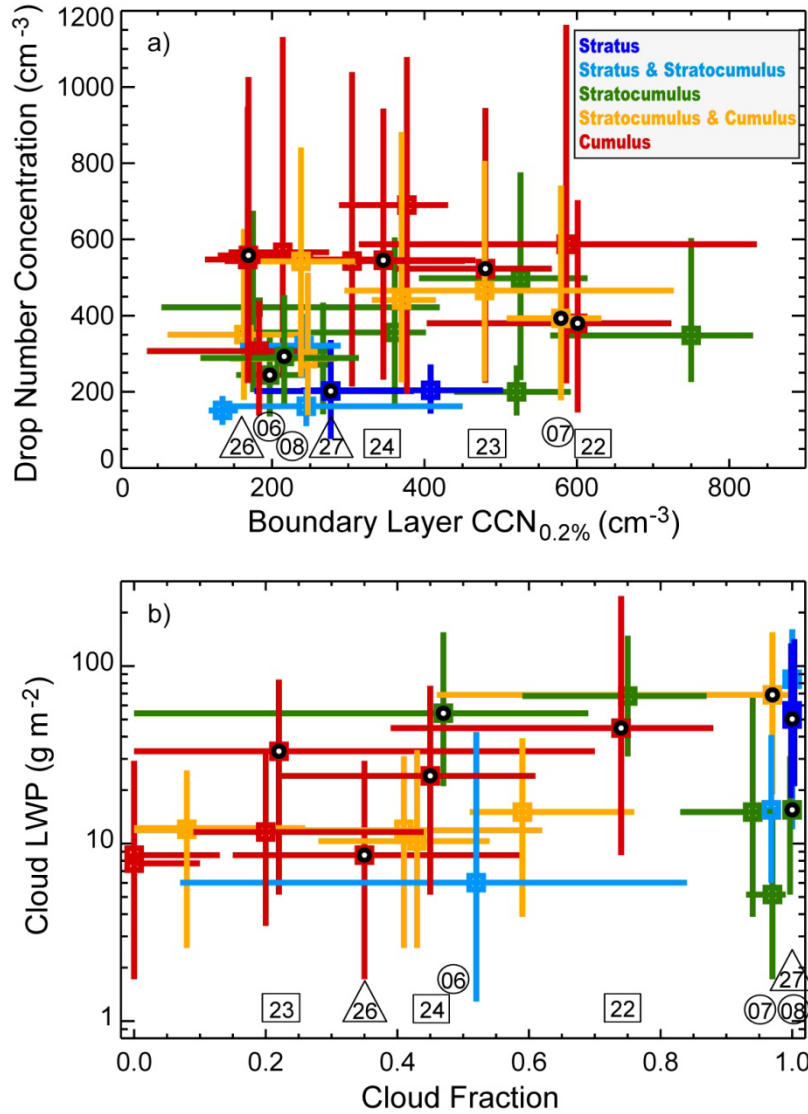


Figure 1. RACORO campaign cloud statistics. Boxes are medians and bars extend between the 25th and 75th percentiles. Colors indicate the dominant cloud type as given in the legend (from *Vogelmann et al.* [2012]). The selected days are denoted by the dates (in May) given below their medians (black-filled squares), and the case number is denoted by the symbols encircling the date, where box=case 1 (all Cu), triangle=case 2 (Cu and St), circle=case 3 (Sc, Sc-Cu and St). (a) Microscale properties of cloud-droplet number concentration (N_d) and CCN at 0.2% supersaturation, $\text{CCN}_{0.2\%}$. $\text{CCN}_{0.2\%}$ are from all on-station measurements outside of cloud ($\text{LWC} < 0.01 \text{ g m}^{-3}$), which are concentrated at altitudes near the cloud layer. (b) Macroscale properties of cloud fraction and cloud LWP. Cloud fraction is from the ARSCL cloud mask of overhead cloud frequency using a 20-min moving average. LWP is the product of cloud thickness, from ARSCL, and LWC, from in-situ CAS measurements (see text). The LWP median and percentiles are products of the respective LWC and cloud thickness percentiles. Eight flights are not plotted in (b) because they were not coincident with the ARSCL measurements.

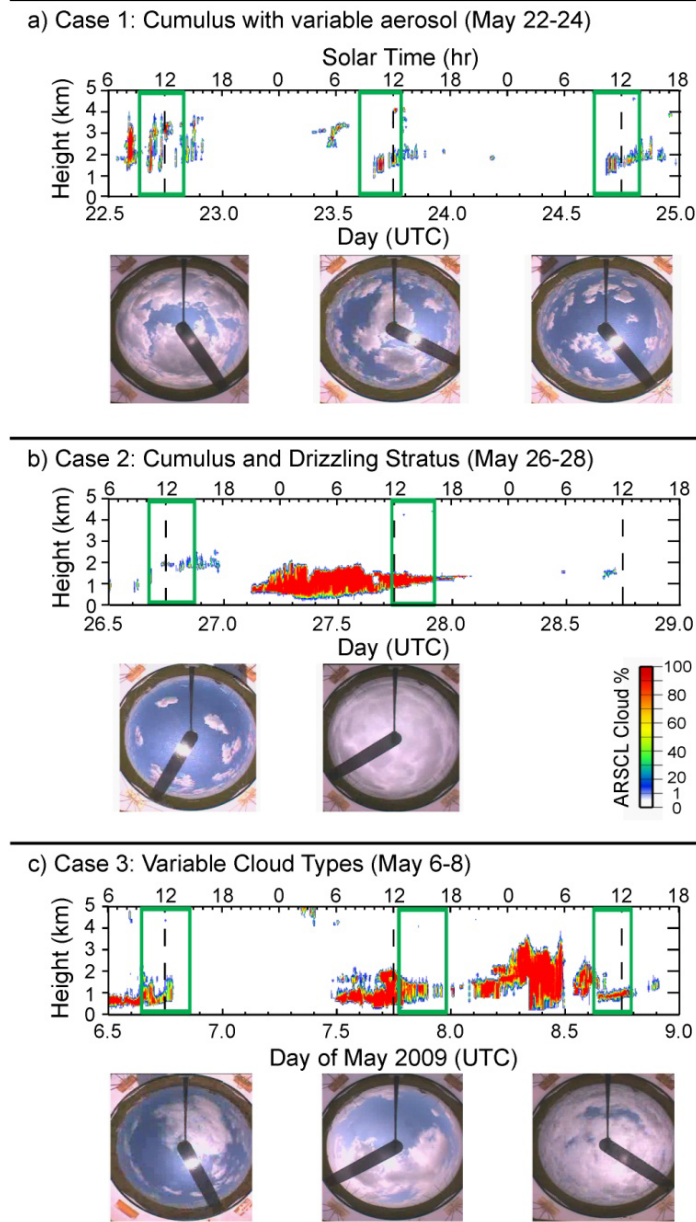
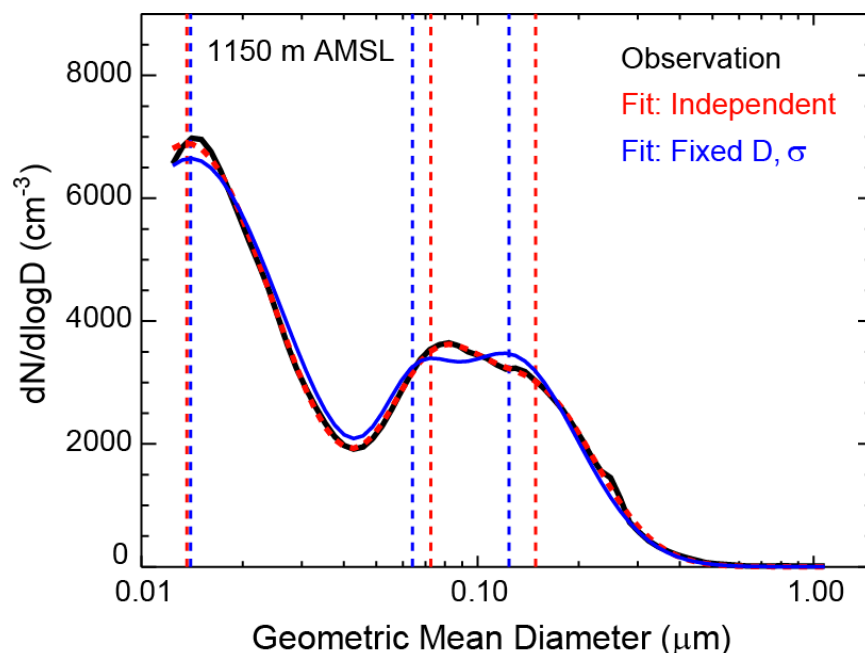


Figure 2. Three RACORO case study periods. Shown are the 60-h case study periods: a) case 1 cumulus, 22-24 May; b) case 2 cumulus and drizzling stratus followed by clear sky, 26-28 May; and c) variable cloud types, 6-8 May (see text for details). The bottom axis is time as decimal day in May (UTC) and the top axis is local solar time (UTC minus 6 h); vertical dashed lines indicate solar noon. The time-height ARSCL cloud fractional occurrence per 10-min interval is given for each case, and flight periods are indicated by green boxes. Below each period is a representative hemispheric view of the cloud field from the Total-Sky Imager (TSI) at approximately the middle of the flight. The radial band in the TSI images is the sun-blocking band. No flight was flown on the last day of case 2, 28 May. Note the enhanced resolution at the bottom of the ARSCL color scale.

1716



1717

1718 **Figure 3. Representative example of three-mode aerosol size distribution fits.** Data are
 1719 from the second spiral on 23 May for the 100-m height-interval centered on 1,150 m above mean
 1720 sea-level (AMSL), or about 800 m above ground level. The black line is the combined average
 1721 of the SMPS and PCASP observations. The red dashed curve is the best fit obtained by allowing
 1722 the geometric mean diameter, D , and geometric standard deviation, σ , to vary independently with
 1723 altitude bin (D values indicated by vertical red lines). The blue curve is the resulting fit when D
 1724 and σ are fixed per spiral profile and the number concentrations per mode are refit per height
 1725 interval (fixed D values indicated by the vertical blue lines). The quality of the fit degrades
 1726 slightly when D and σ are fixed but it still provides good overall agreement, particularly for the
 1727 particle diameters greater than $\sim 0.1 \mu\text{m}$ that are most active as CCN for supersaturations $\leq 0.6\%$
 1728 in our case studies.

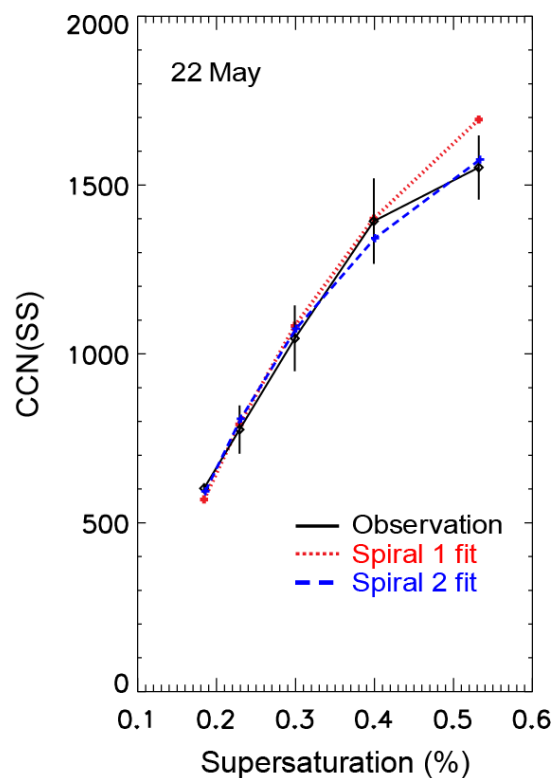


Figure 4. Example of κ fitting. The observed CCN supersaturation spectrum, CCN(SS), is the average of the observations (i.e., averages of the 60 s 1-Hz averages) from 22 May (solid black line), and standard errors are given for each supersaturation (0.19, 0.23, 0.30, 0.40, 0.53). The same averaged CCN data are used in separate κ analyses with the aerosol fits from the spiral at the start (spiral 1) and end of the flight (spiral 2). The κ values determined are consistent between the spirals being, respectively, 0.08 and 0.07.

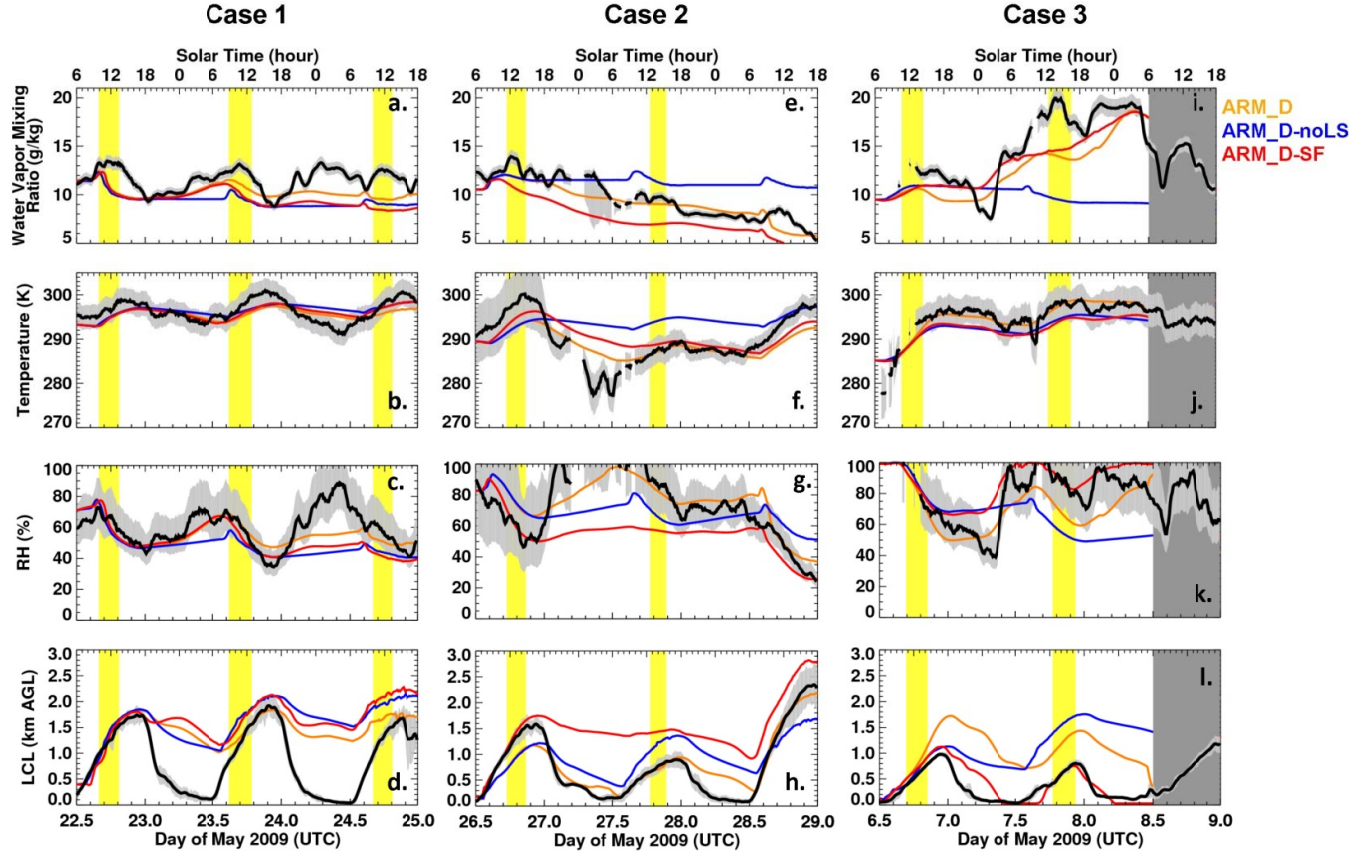


Figure 5. Sensitivity to advection and surface forcing components. Observations and DHARMA trial simulations are shown for the three case study periods in terms of lifting condensation level height (Z_{LCL}) and mid-boundary layer average (300-500 m) water vapor mixing ratio ($\overline{q_v}$), temperature (\overline{T}), and relative humidity (\overline{RH}). Local solar time (LST) is indicated at the top of the figure and decimal day (UTC) in May at the bottom. A one-hour moving average is applied to the observations, as described in Appendix B. Observations and their uncertainties are represented by, respectively, black lines and gray shading. Simulations are for the ARM standard domain without relaxation (ARM_D, orange), for surface forcing as in ARM_D but with no large-scale forcing (ARM_D-noLS, blue), and large-scale forcing as in ARM_D but with the domain-averaged EBBR surface fluxes replaced by ECOR values at the SGP central facility (ARM_D-SF, red). Yellow vertical bars indicate the aircraft flight periods. Simulations are not plotted for case 3 after 8.5 May, as discussed in the text.

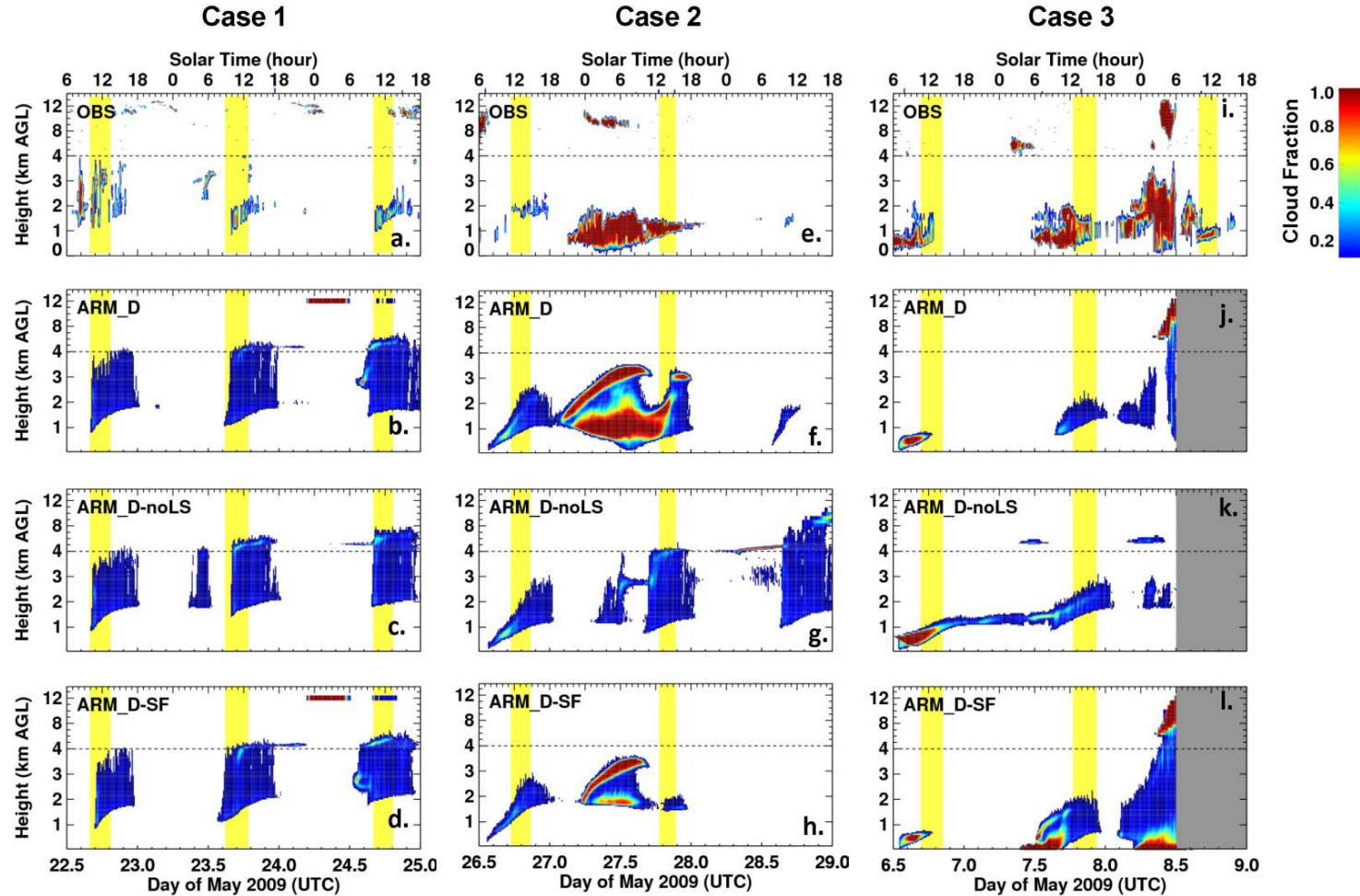


Figure 6. Cloud fraction profile sensitivity. Shown are cloud fraction profiles observed by ARSCL and from simulations for the three case studies periods for ARM_D, ARM_D-SF, and ARM_D-noLS. Local solar time (LST) is indicated at the top and decimal day (UTC) in May at the bottom. A two-toned vertical scale is used, where the vertical region below 4 km is expanded and above 4 km is reduced; the partition between the regions is indicated by a dashed line. This enables viewing the full tropospheric cloud profiles without sacrificing details of the boundary layer clouds, as would be the case if a single linear scale were used. Yellow vertical bars indicate the aircraft flight periods. Simulations are not plotted for case 3 after 8.5 May, as discussed in the text.

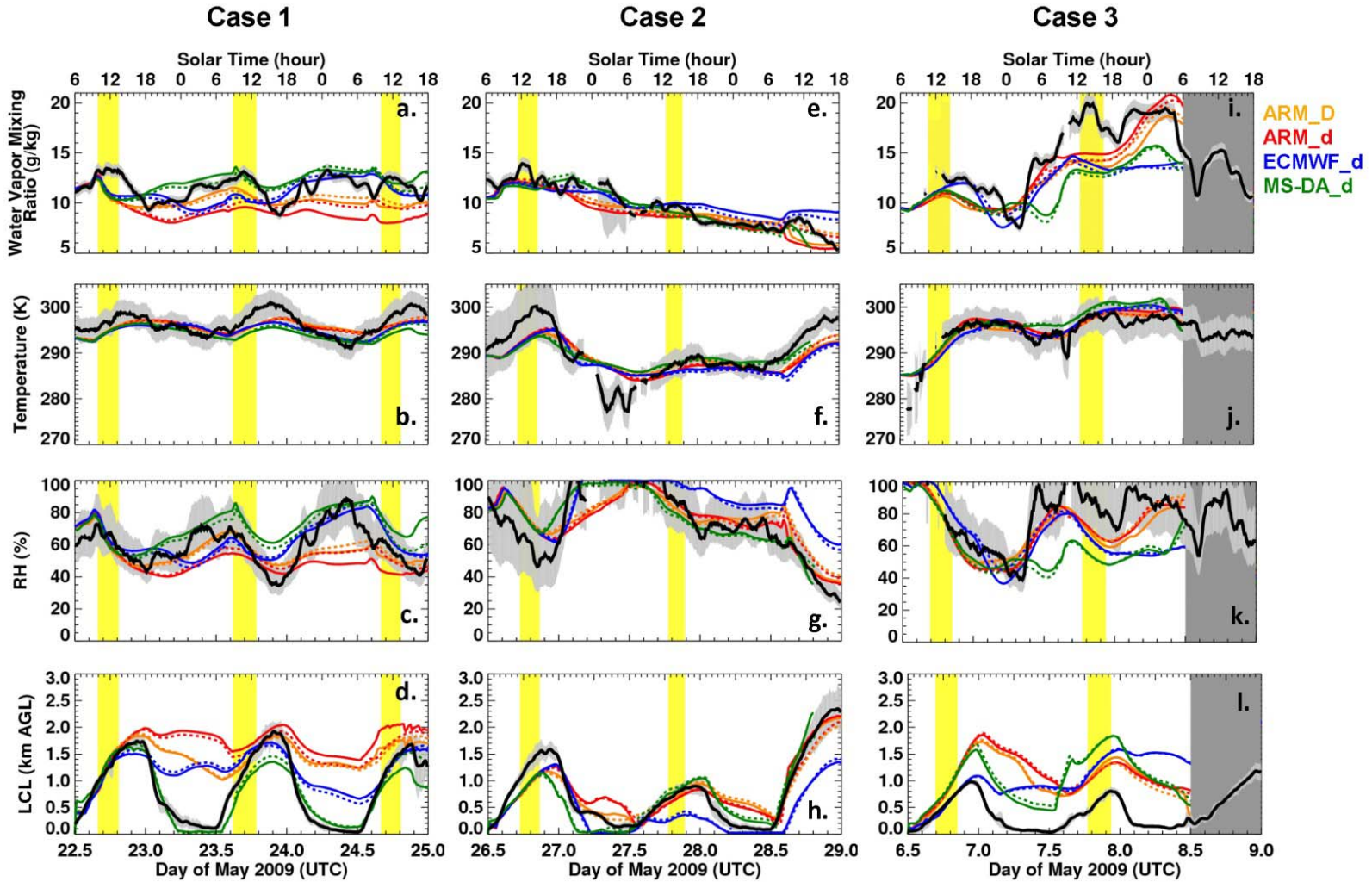


Figure 7. Simulated atmospheric state for ensemble forcings. As for Figure 5, but for DHARMA simulations using the ARM forcing for the standard 300-km domain (ARM_D, orange) and 150-km reduced domain (ARM_d, red), and for the ECMWF reduced domain (ECMWF_d, blue). Reduced-domain MS-DA simulations (MS-DA_d, green) are run using WRF (see text). Dashed lines use 12-h thermodynamic relaxation, and solid lines are without relaxation. Yellow vertical bars indicate the aircraft flight periods. (Note that case 2 simulations end at 28.8 May due to WRF stability issues.)

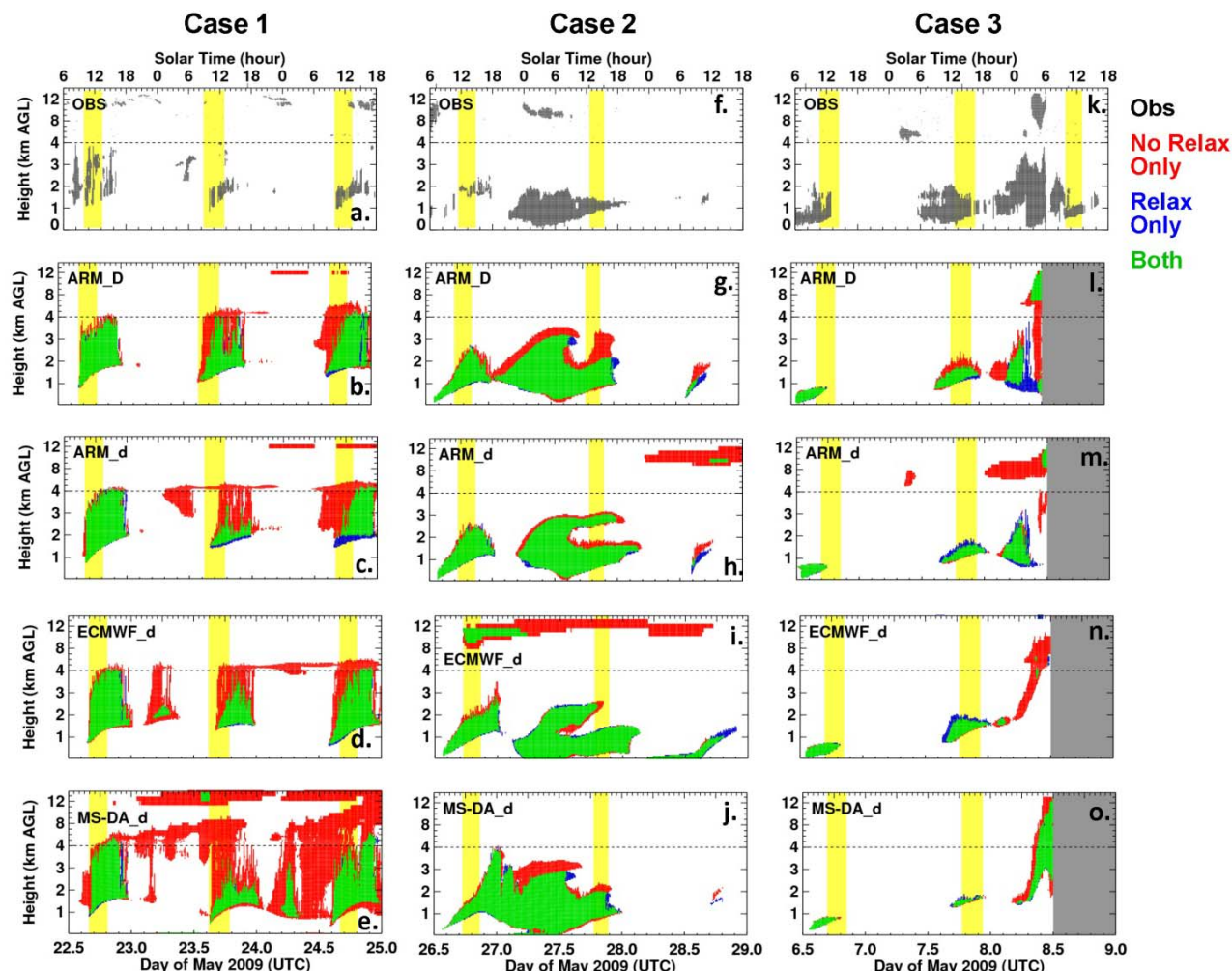


Figure 8. Cloud-mask profile for ensemble forcings. 2D cloud masks show the time-height location of cloud (not cloud fraction) for ARSCL observations and DHARMA simulations with and without 12-h thermodynamic relaxation for ARM_D, ARM_d, and ECMWF_d. The MS-DA_d simulations are run using WRF. Green indicates where the simulations with and without relaxation both have cloud, red is where cloud is only present without relaxation, and blue is where cloud is only present with relaxation. The same two-toned vertical scale is used as in Fig. 6, where the vertical region below 4 km is expanded and above 4 km is reduced. Yellow vertical bars indicate the aircraft flight periods.

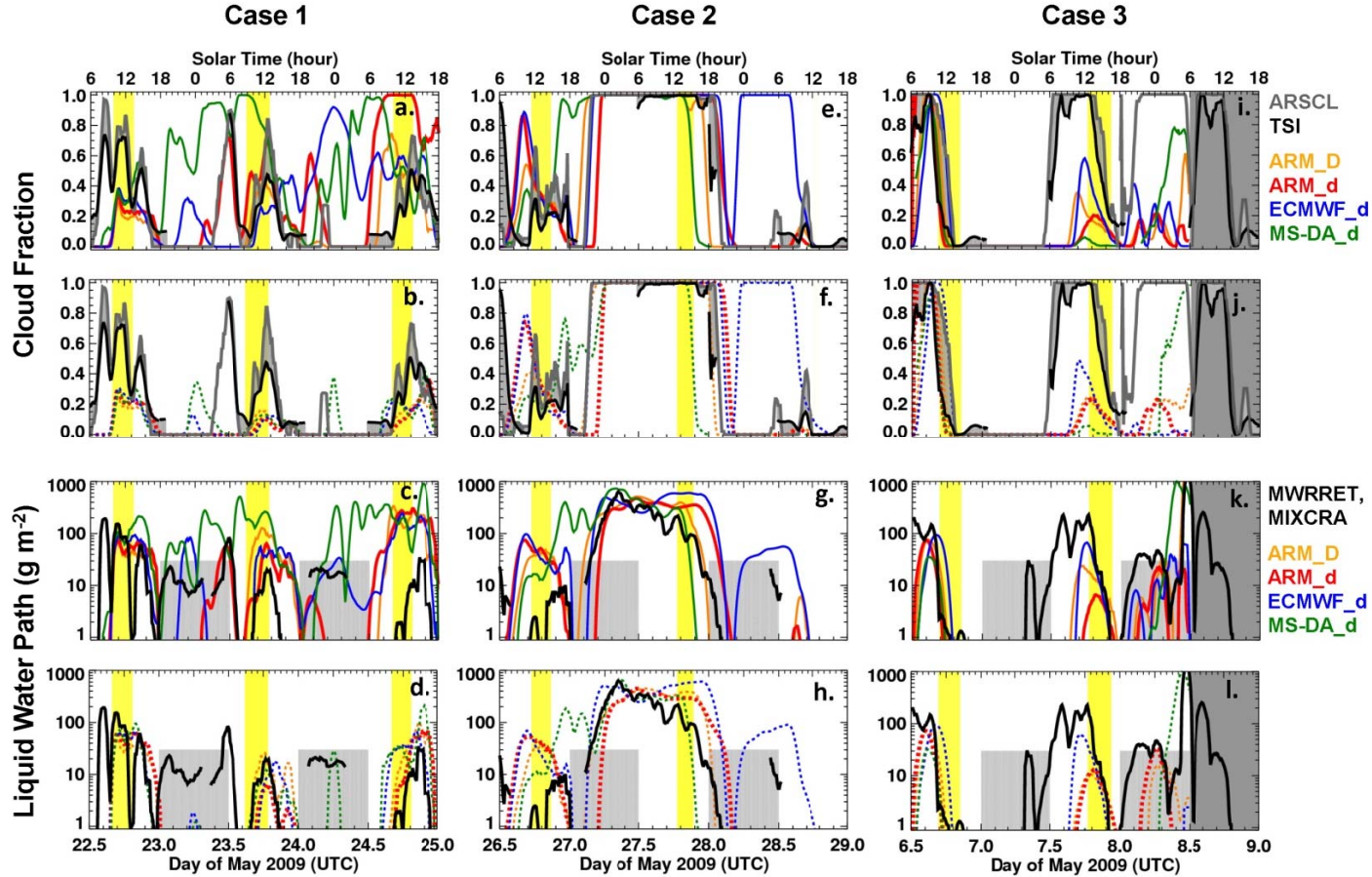
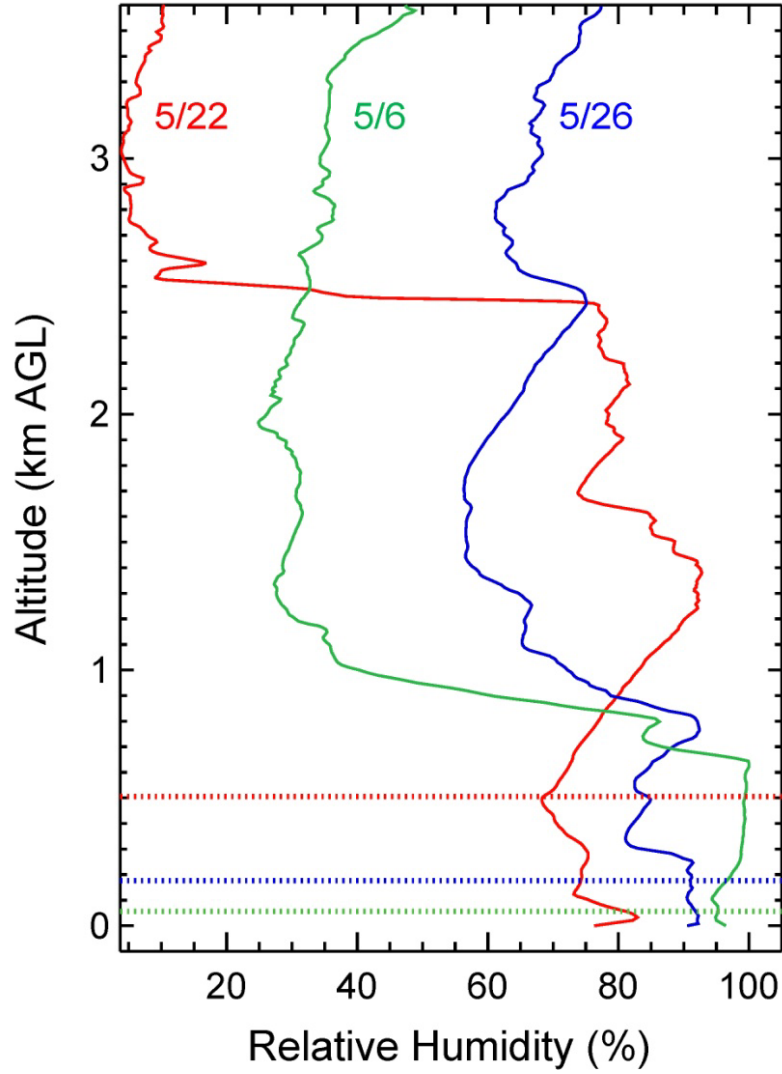


Figure 9. Cloud fraction and $\overline{\text{LWP}}$ for ensemble forcings. Observed and simulated domain-averaged cloud fraction and $\overline{\text{LWP}}$ for the three cases (columns) using DHARMA for ARM_D, ARM_d, ECMWF_d, and WRF for MS-DA_d. Shown are cloud fraction without thermodynamic relaxation (row 1), cloud fraction with relaxation (row 2), $\overline{\text{LWP}}$ without relaxation (row 3), and $\overline{\text{LWP}}$ with relaxation (row 4). Cloud fraction observations are from the TSI (black line) and ARSCL (dark gray line) with shading in between when both are present. $\overline{\text{LWP}}$ observations are from MIXCRA during the daytime and MWRRet during the nighttime; when MWRRet is used its uncertainty is indicated by gray shading below 30 g m^{-2} . Observed cloud fraction and $\overline{\text{LWP}}$ use one-hour smoothing for clarity. Yellow vertical bars indicate the aircraft flight periods. Note that simulated cloud fraction and $\overline{\text{LWP}}$ are computed for cloud $< 8 \text{ km}$.



1777

1778 **Figure 10. Initial condition RH profiles.** The lower-atmospheric RH profiles (< 3.5 km) are
 1779 shown from the observed soundings used as initial conditions for the three cases: case 1
 1780 sounding for 22 May 11:30 UTC (red), case 2 sounding for 26 May 11:28 UTC (blue), and case
 1781 3 sounding for 6 May 11:27 UTC (green). Their respective Z_{LCL} values are given as horizontal
 1782 dashed lines.

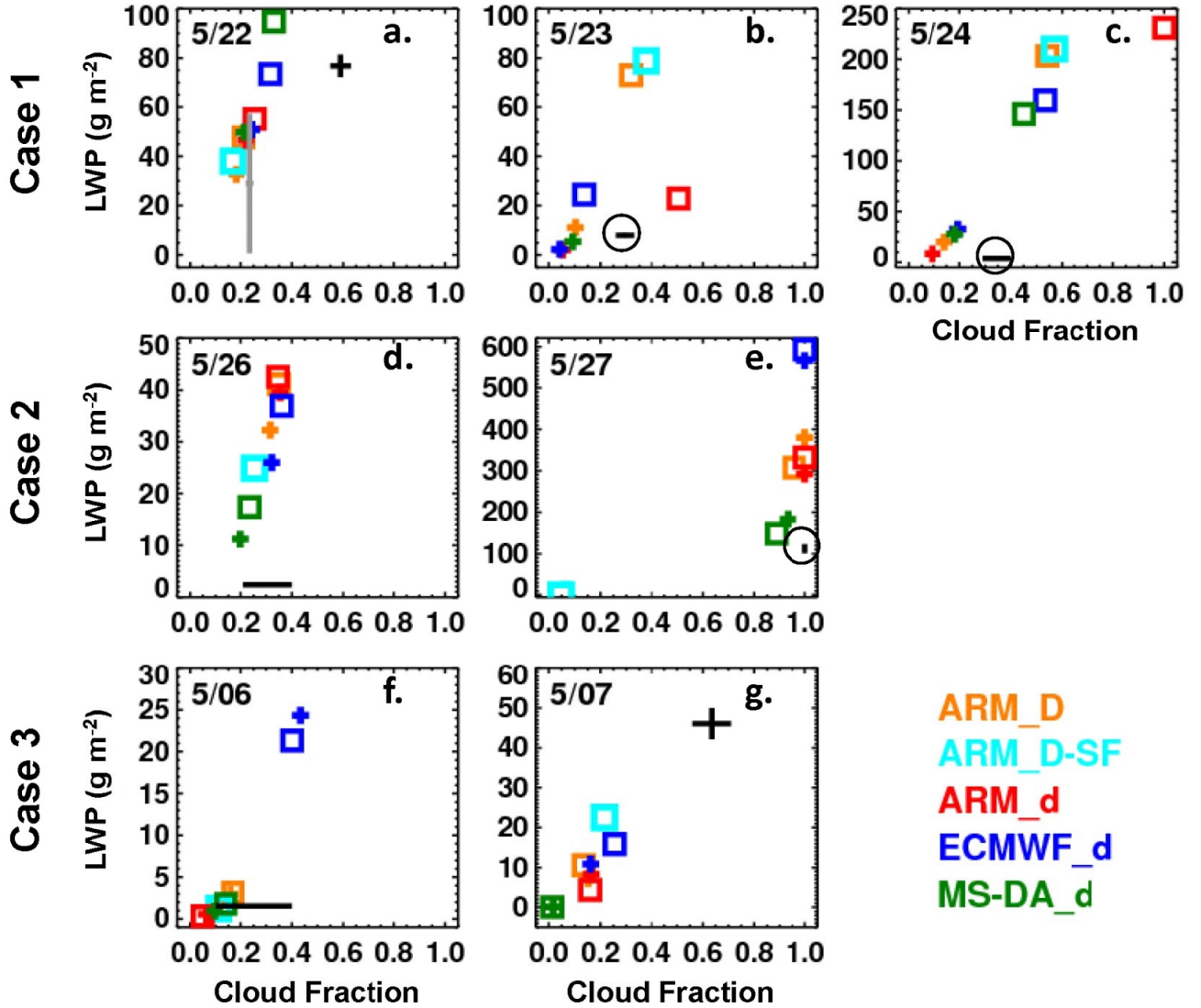
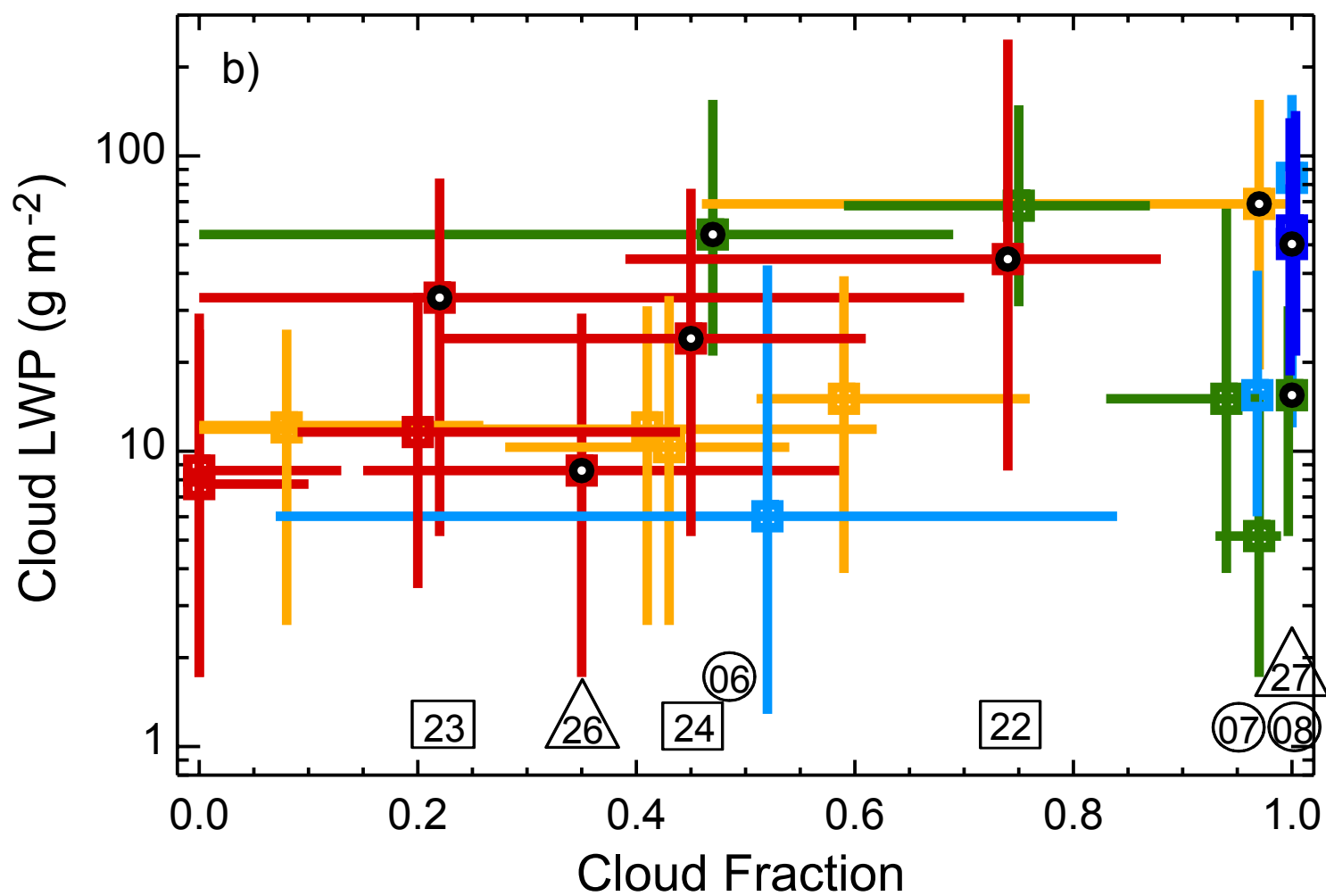
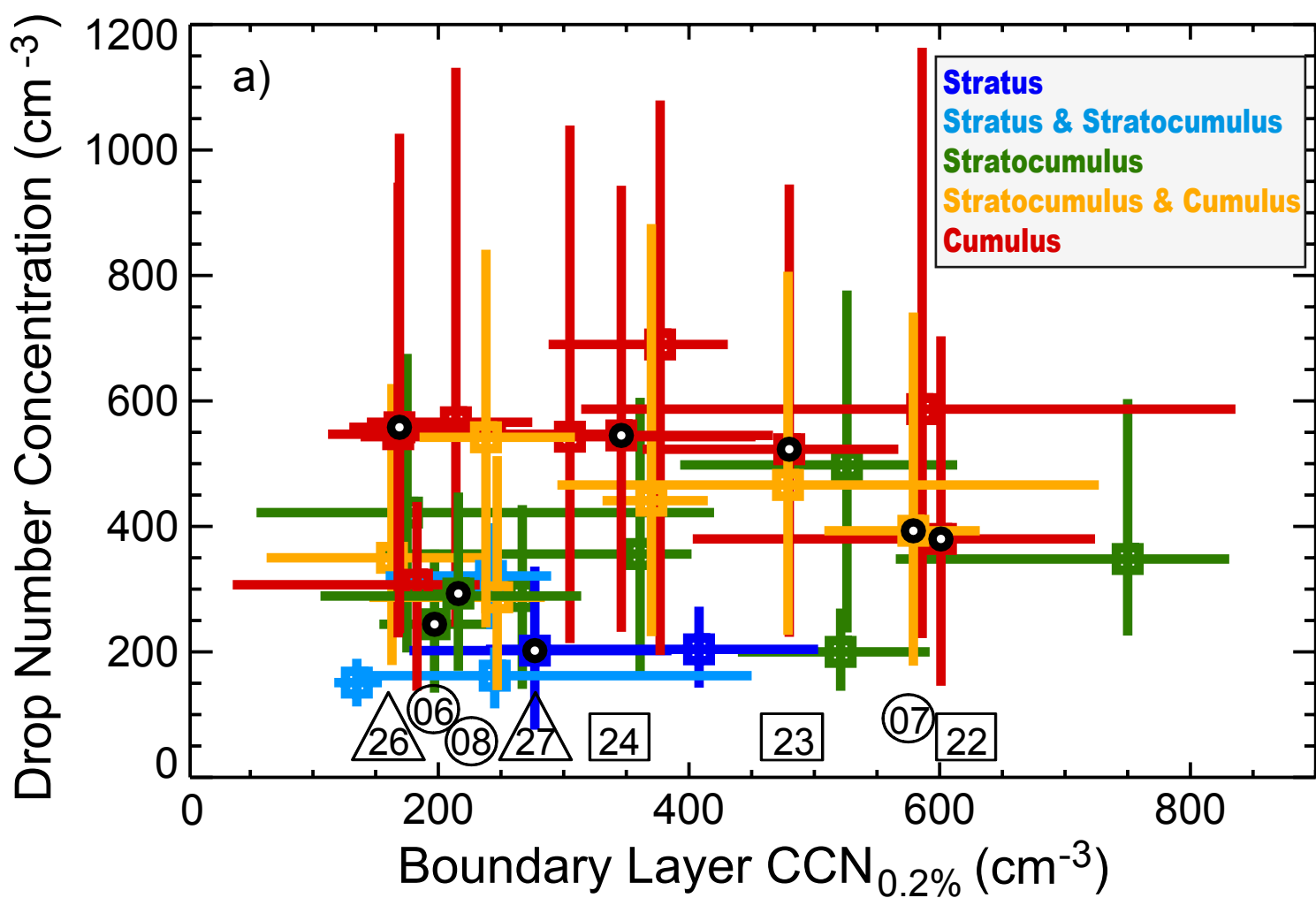
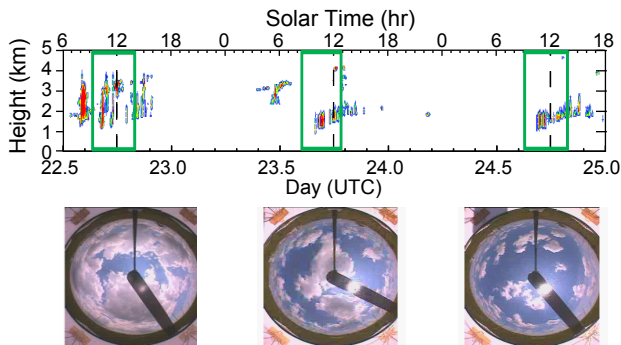


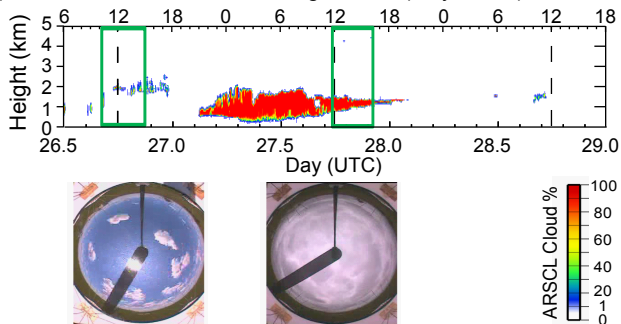
Figure 11. Flight period relationship between mean cloud fraction and $\overline{\text{LWP}}$. Flights for the three cases are given per row. A point is an average over each flight period of the domain cloud fraction and domain-mean $\overline{\text{LWP}}$ computed from the simulations (see legend); pluses are with thermodynamic relaxation and squares are without it. The horizontal extent of the black observation bars represents the cloud fraction range from the TSI and ARSCL estimates using a 20-min moving average (as in Fig. 1b). The vertical bar—when visible—is the MIXCRA retrieval uncertainty after averaging across the domain, where the cloud-free measurement uncertainty is effectively zero. Observations are circled for clarity in panels b, c, and e. The gray bar in 5/22 estimates the values excluding the early morning residual layer (see text). Results for the 5/8 flight are not plotted because most simulations are invalid after an earlier passage of a squall line to the north of the central facility (see text). MS-DA without relaxation is not plotted in (b) because it is significantly off scale (cloud fraction = 0.87, $\overline{\text{LWP}}$ = 199 g m⁻²).



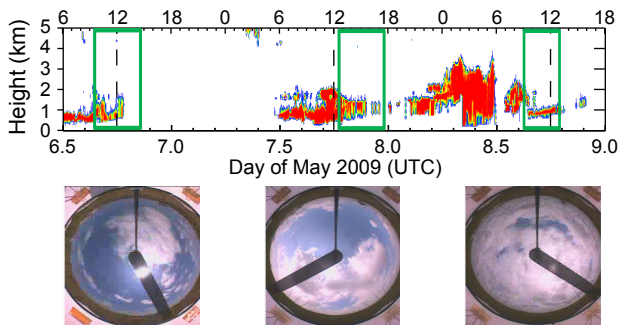
a) Case 1: Cumulus with variable aerosol (May 22-24)

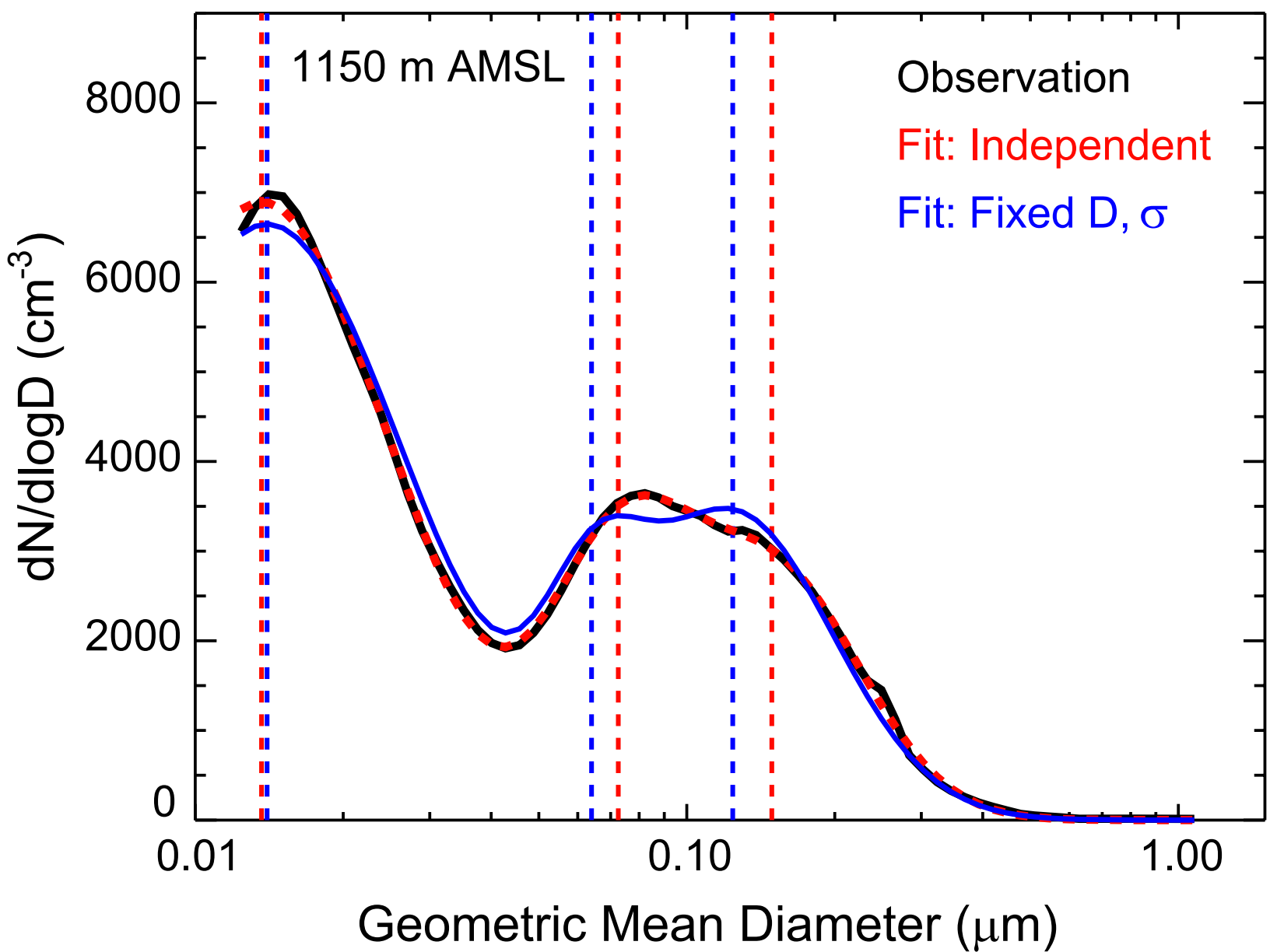


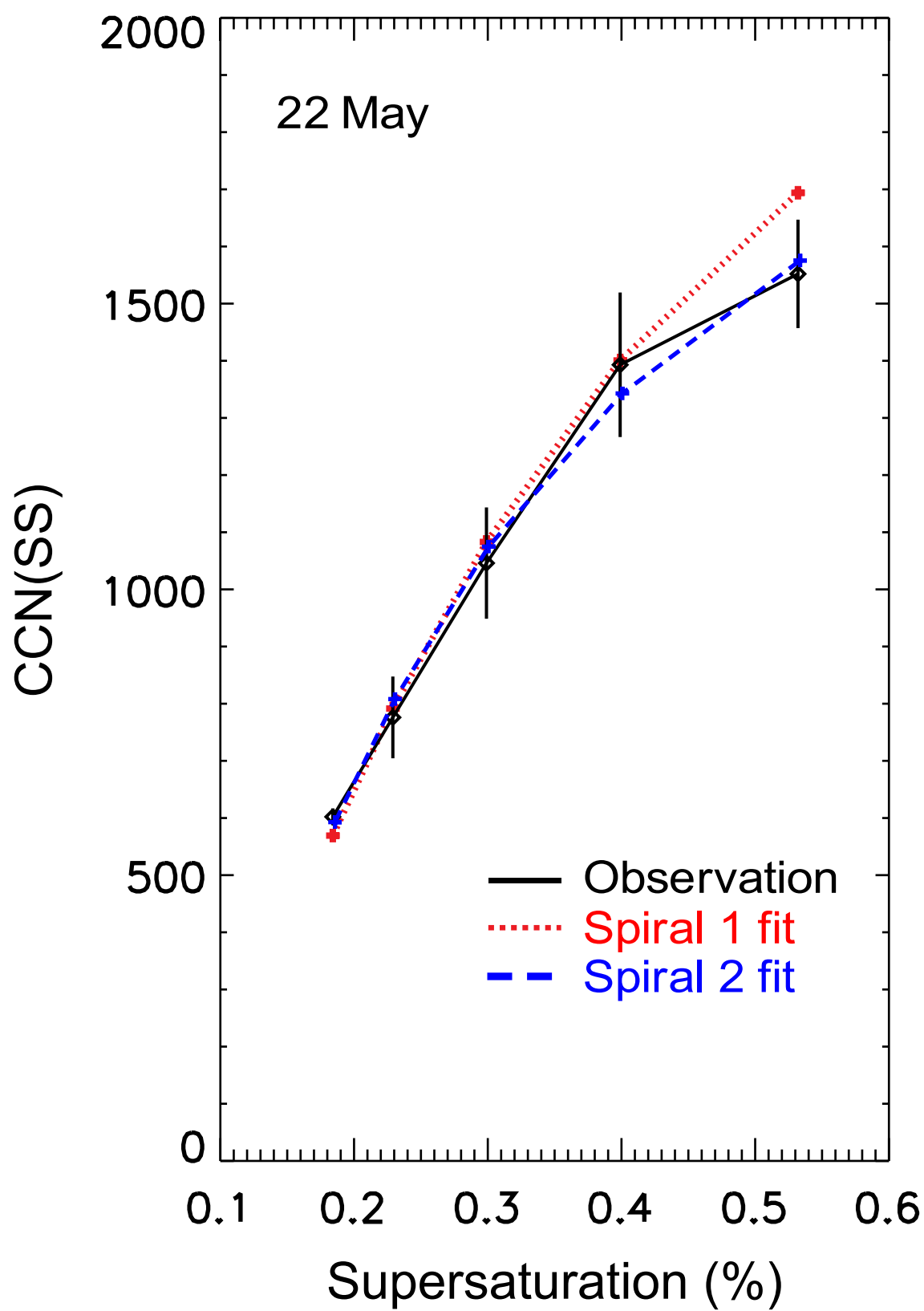
b) Case 2: Cumulus and Drizzling Stratus (May 26-28)



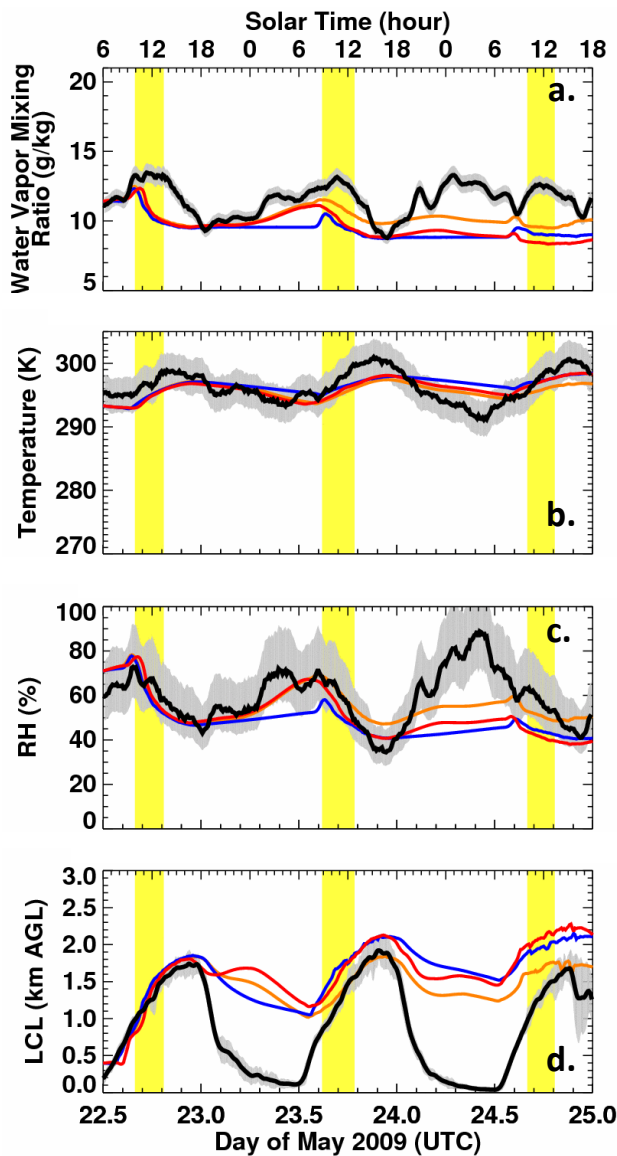
c) Case 3: Variable Cloud Types (May 6-8)



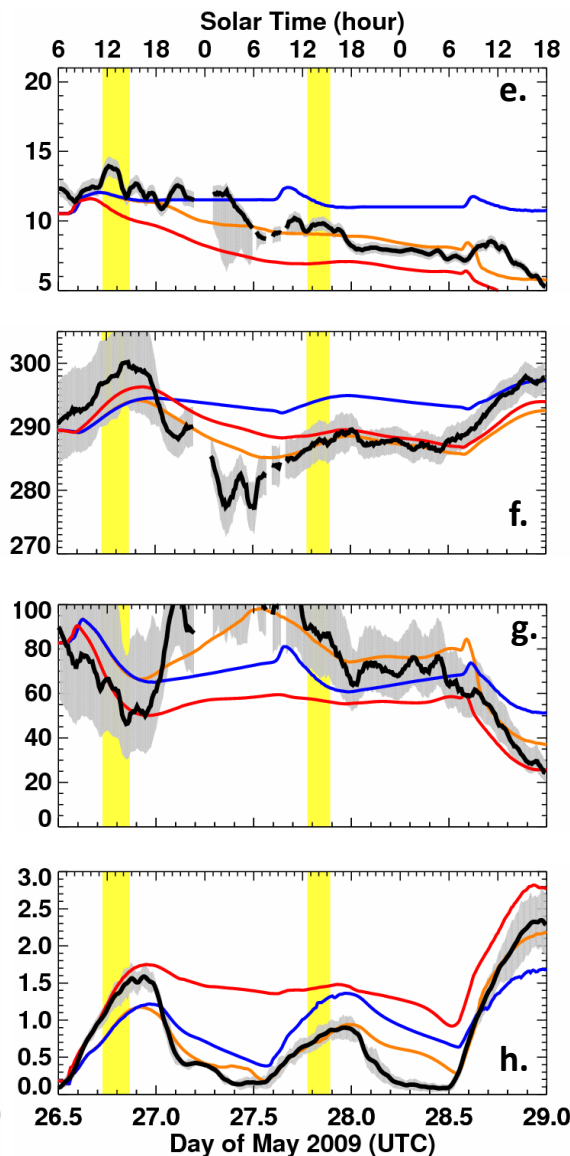




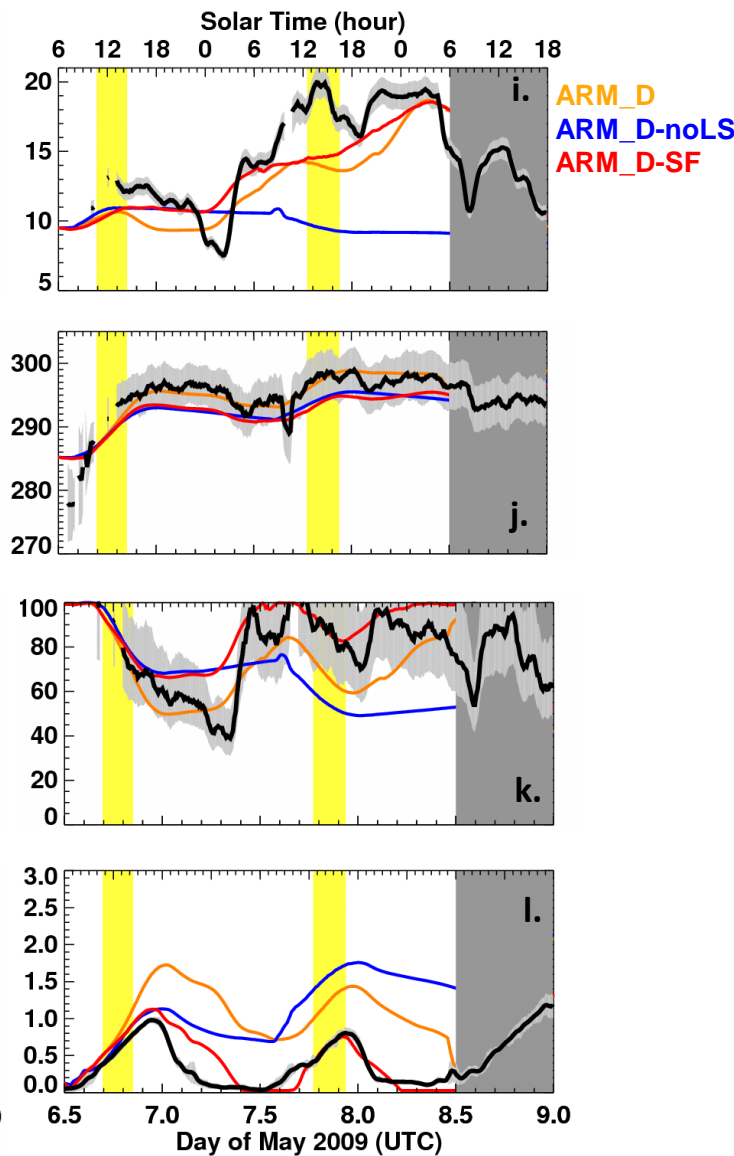
Case 1



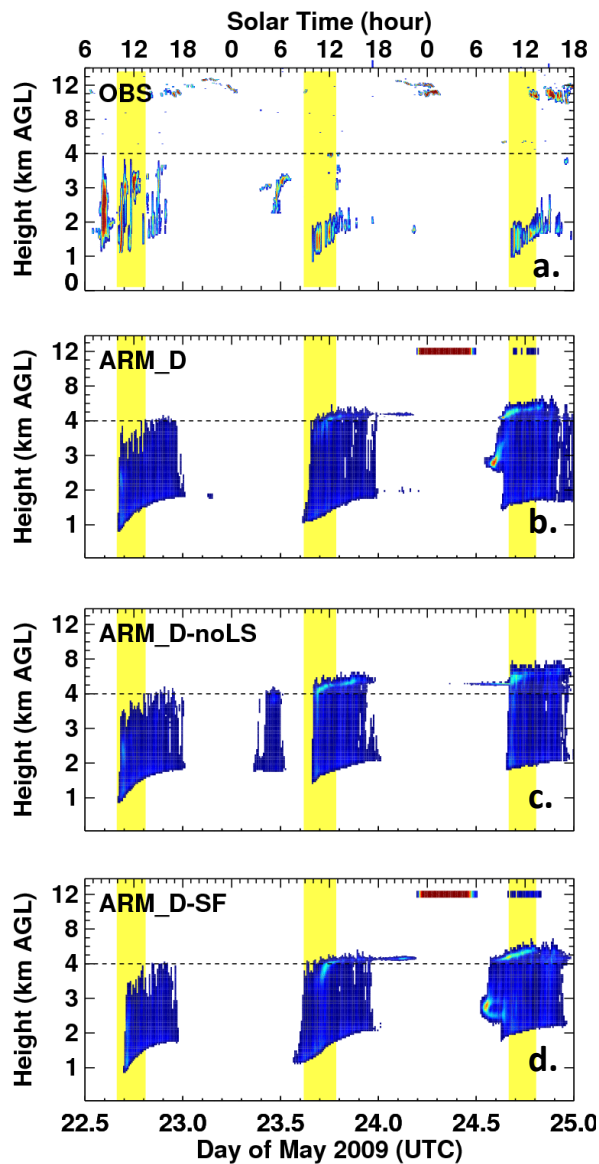
Case 2



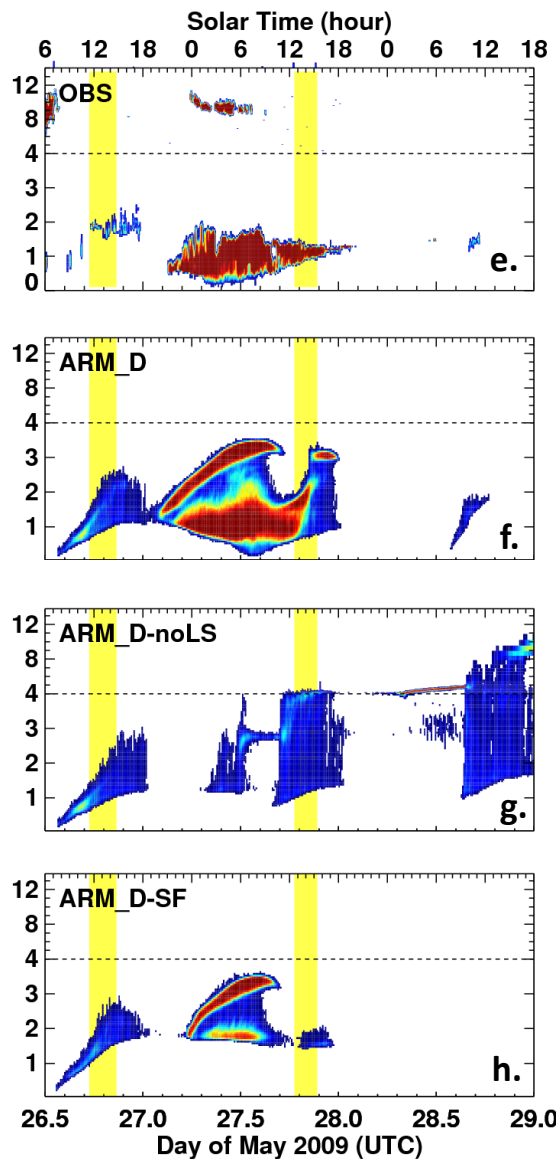
Case 3



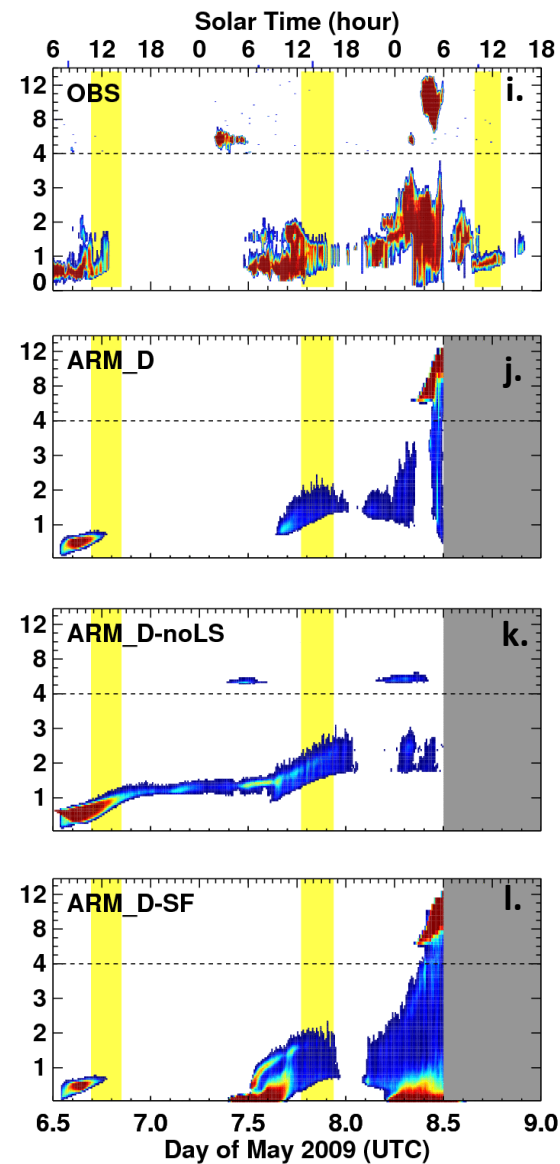
Case 1

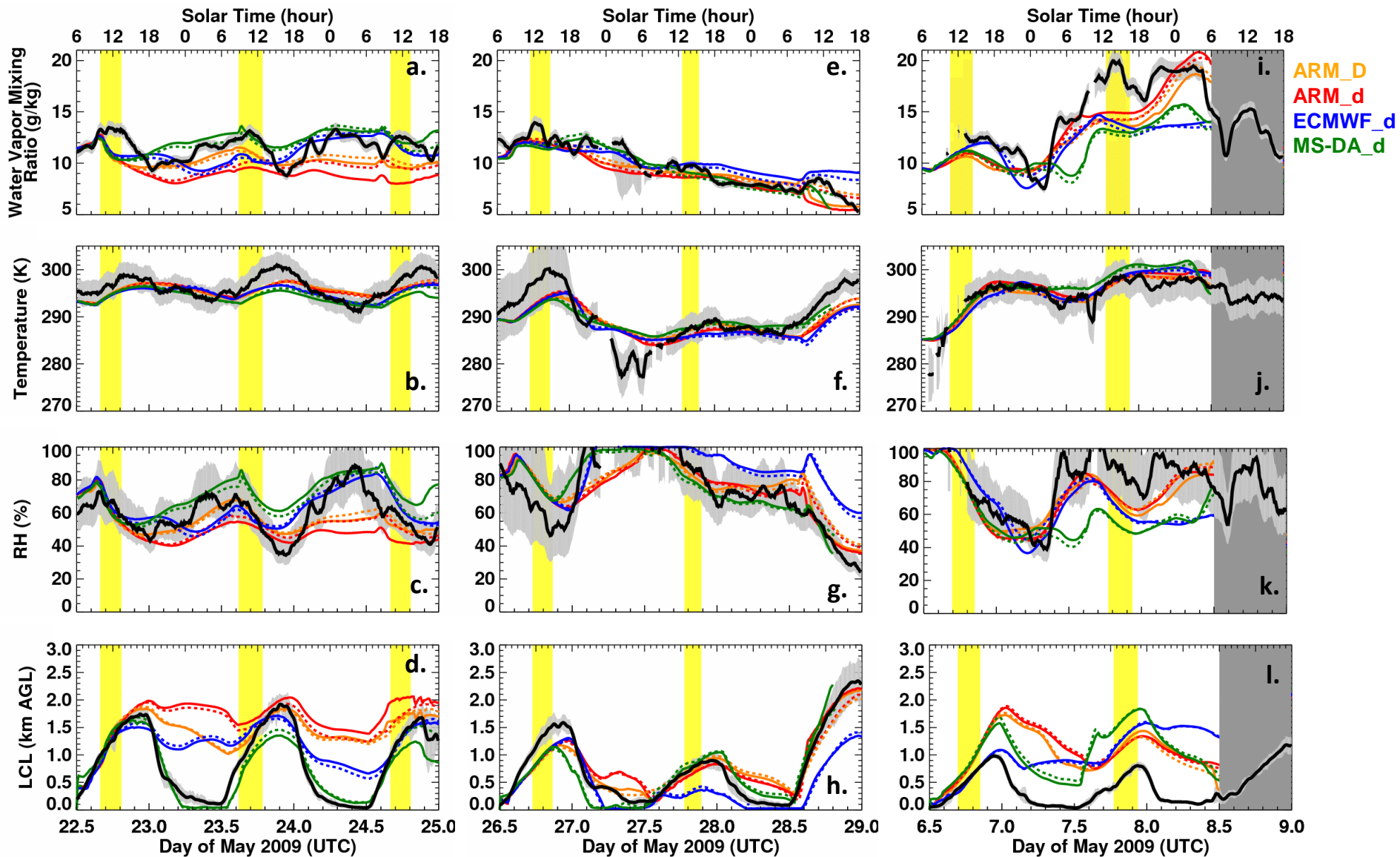


Case 2

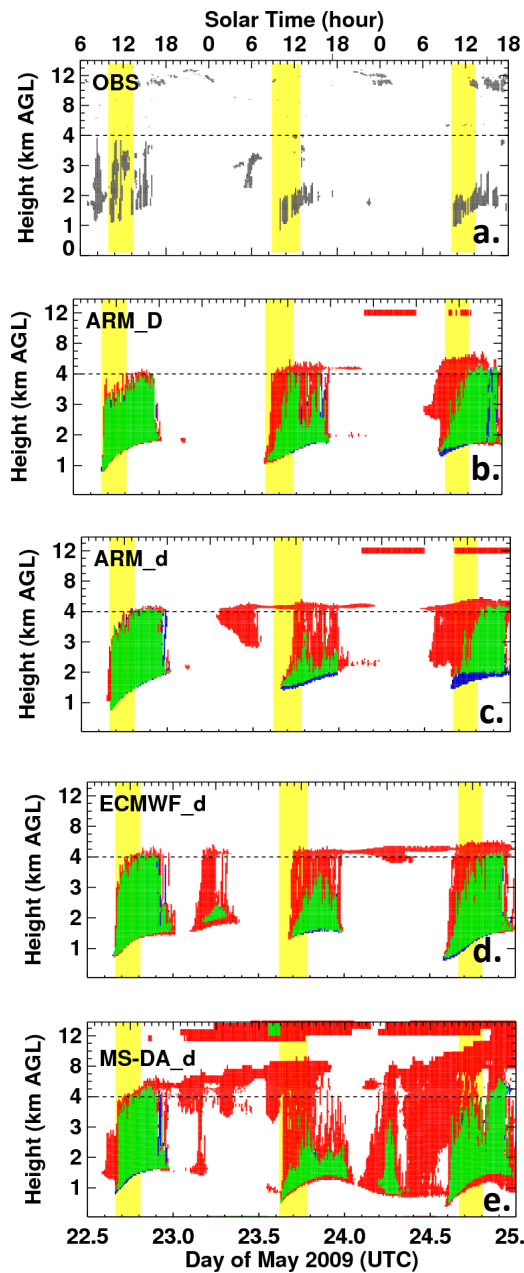


Case 3

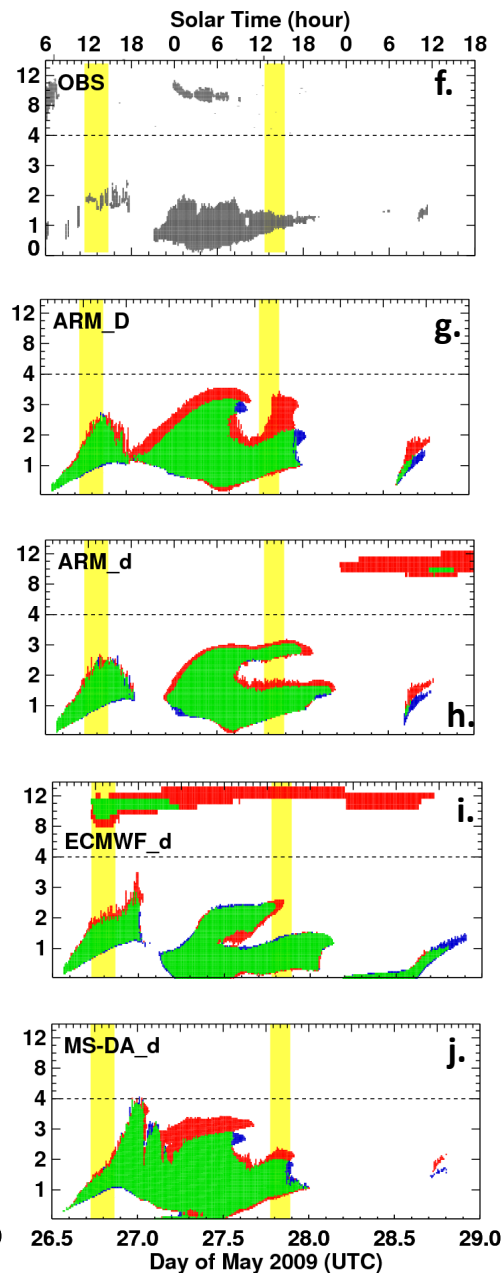


Case 1**Case 2****Case 3**

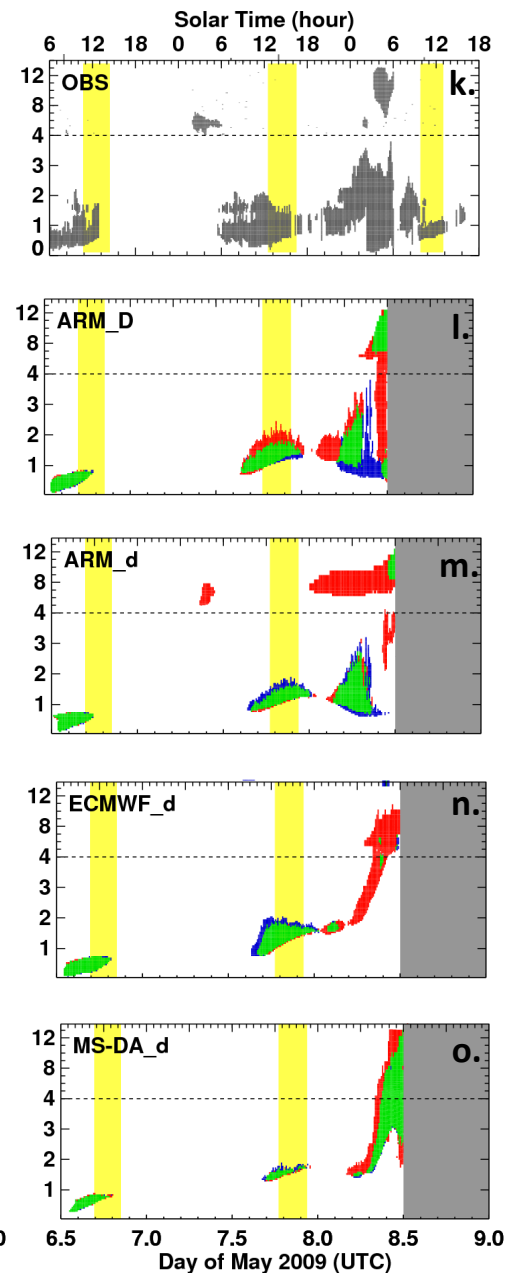
Case 1



Case 2



Case 3



Obs

No Relax
Only

Relax
Only

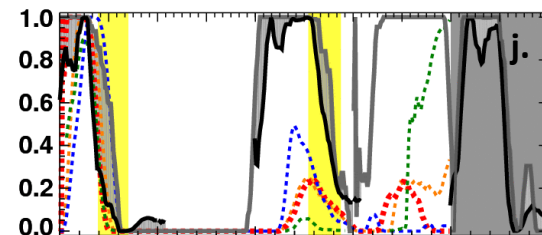
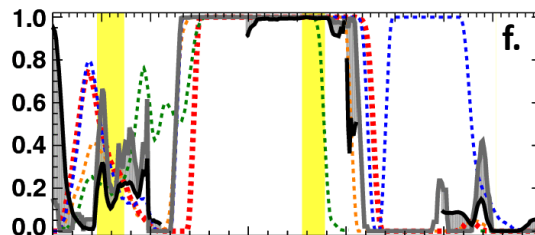
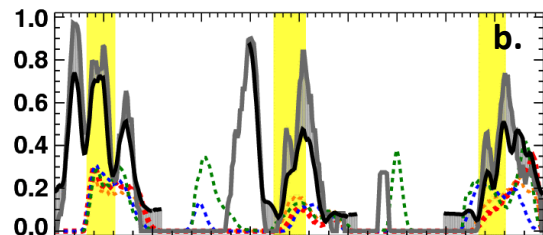
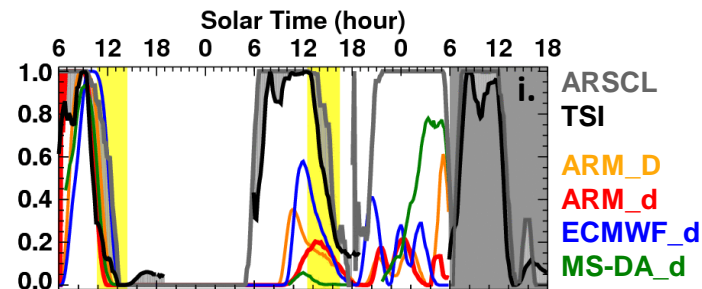
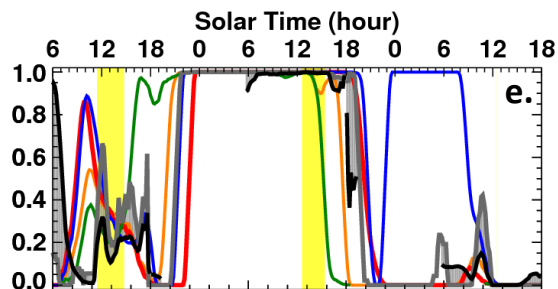
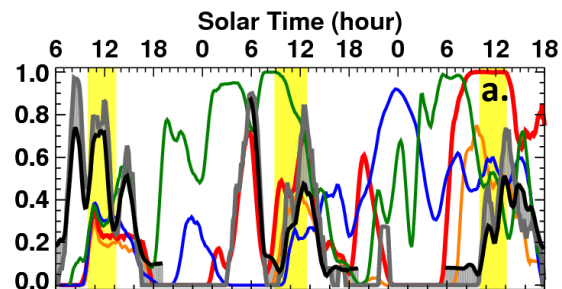
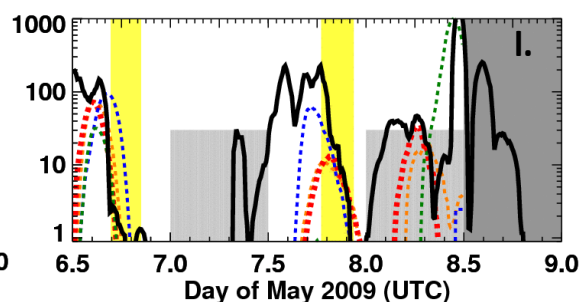
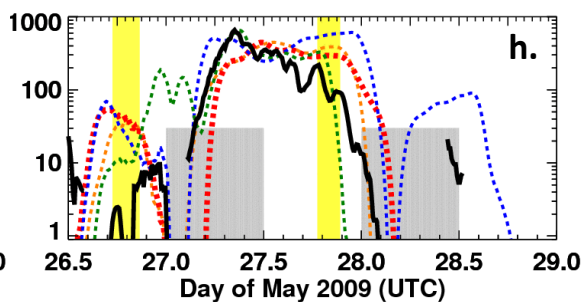
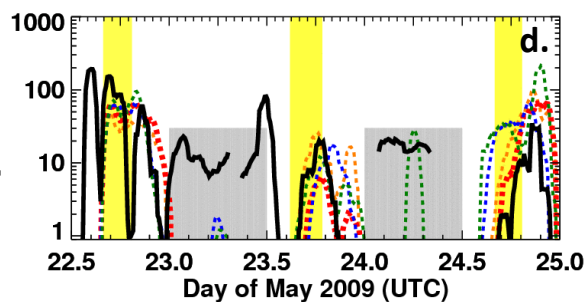
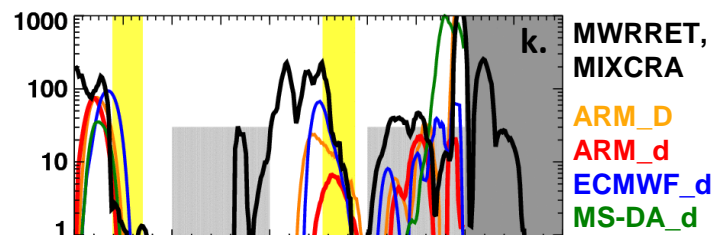
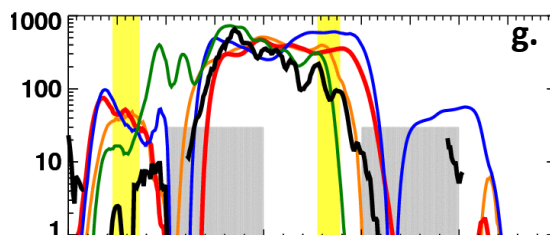
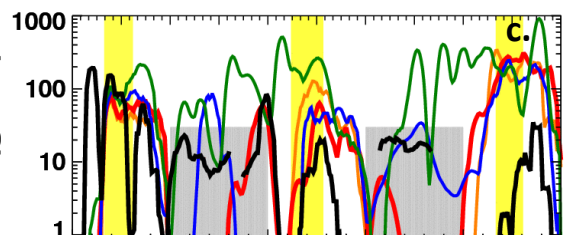
Both

Case 1

Case 2

Case 3

Cloud Fraction

Liquid Water Path (g m^{-2})

ARSCL

TSI

ARM_D

ARM_d

ECMWF_d

MS-DA_d

MWRRET,
MIXCRA

ARM_D

ARM_d

ECMWF_d

MS-DA_d

